| | |
|---|---|
| Project acronym: | RECODE |
| Project title: | Policy RECommendations for Open access to research Data in Europe |
| Grant number: | 321463 |
| Programme: | Seventh Framework Programme for Science in Society |
| Objective: | SiS-2012.1.3.3-1: Scientific data: open access, dissemination, preservation and use |
| Contract type: | Co-ordination and Support Action |
| Start date of project: | 01 February 2013 |
| Duration: | 24 months |
| Website: | www.recodeproject.eu |

# Deliverable D5.1:
# Policy guidelines for open access and data dissemination and preservation


| | |
|---|---|
| Author(s): | Victoria Tsoukala, Marina Angelaki and Vasso Kalaitzi (National Documentation Centre); Bridgette Wessels, Lada Price, Mark J. Taylor and Rod Smallwood (University of Sheffield); Peter Linde (Blekinge Institute of Technology); Jeroen Sondervan (Amsterdam University Press); Susan Reilly (LIBER); Merel Noorman and Sally Wyatt (Royal Netherlands Academy of Arts and Sciences); Lorenzo Bigagli (National Research Council of Italy); Rachel Finn, Thordis Sveinsdottir and Kush Wadhwa (Trilateral Research and Consulting) |



| | |
|---|---|
| Dissemination level: | Public |
| Deliverable type: | PUBLIC |
| Version: | FINAL (2/02/2015) |
| Submission date: | Due:January 2015-Submitted: February 2015 |

## CONTENTS

## LIST OF ACRONYMS

ANDS: Australian National Data Service
APARSEN: Alliance for Permanent Access
BD2K: Big Data to Knowledge
CERN: European Organization for Nuclear Research
DANS: Dutch Archiving and Networking Services
DCC: Digital Curation Centre
DFG: Deutsche Forschungsgemeinschaft
DMP: Data Management Plan
DOI: Digital Object Identifier
DSA: Data Seal of Approval
DUI: Data Usage Index
EASY: Electronic Archiving System
EC: European Commission
EPSRC: Engineering and Physical Sciences Research Council
ESRC: Economic and Social Research Council
GEOSS: Global Earth Observation System of Systems
HEFCE: Higher Education Funding Council for England
INSPIRE: Infrastructure for Spatial Information in the European Community
IODE: International Oceanographic Data and Information Exchange
IPR: Intellectual Property Right
JISC: Joint Information Systems Committee
JoRD: Journal Research Data Policies
LERU: League of European Research Universities
NEH: National Endowment for the Humanities
NERC: Natural Environment Research Council
NIH: National Institute of Health
NOAA: National Oceanic Atmospheric Administration
NODC: National Oceanographic Data Centre
NSF: National Science Foundation
ODE: Opportunities for Data Exchange
OECD: Organisation for Economic Cooperation and Development
PARSE: Permanent Access to the Records of Science in Europe
PLOS: Public Library of Science
PSI: Public Sector Information
RCUK: Research Councils UK
RDA: Research Data Alliance
RDNL: Research Data Netherlands
REF: Research Excellence Framework
SMEs: Small and Medium Enterprises
SSH: Social Sciences and Humanities
STEM: Science, Technology, Engineering and Mathematics
STM: Scientific, Technical and Medical
TDM: Text and Data Mining
UKDA: UK Data Archive
UKDS: UK Data Service

## EXECUTIVE SUMMARY

The RECODE project identified two overarching issues that are inhibiting take-up of policies related to open access to research data: a lack of a coherent open data ecosystem; and a lack of attention to the specificity of research practices, processes and forms of data collections. Against this background, the report provides policy recommendations on open access to research data targeted at key stakeholders in promoting open access: research funders; data managers; research institutions; and publishers. These recommendations will assist the above stakeholders in furthering the goals of open access to research data in each of their organizations and networks. Recommendations include both overarching and stakeholder-specific ones as suggestions to address and attend to these two issues by building on or learning from existing experience. In doing so, RECODE recognises that in some situations it is appropriate to build consensus and transfer good practice across disciplines and stakeholder groups, while in others it is appropriate to enable and support specific groups to maintain their particularity in relation to disseminating, preserving and re-using research data. In addition, for each stakeholder group good practice examples are provided that can serve as "models" for providing open access to research data. Finally, the recommendations and good practice examples will reduce "costs" associated with providing open access, as stakeholders do not need to develop their own expertise, but can use this information as a foundation to develop their own policies, support actions and initiatives.

Developing open access to research data in a way that is informed by awareness of differences in research practices within and between disciplines and characterised by a partnership approach among key stakeholders helps to ensure engagement from the wide range of research communities and to better embed it within research practice and process. As such, the development of open access to research data needs to be:

- ✓ Characterised by a partnership approach involving the key stakeholders, researchers, and institutions
- ✓ Supported by an integrated institutional and technological data infrastructure
- ✓ Guided by ethical and regulatory frameworks
- ✓ Informed by research practices and processes in different fields

In such a context researchers, as the producers of data and consequently the origin point of the data lifecycle, emerge as an essential component of open access processes whose needs, concerns and interests must be considered in order to work towards a strong open data ecosystem.

The policy recommendations presented here are informed by these overall findings as well as a series of horizontal analyses of the RECODE case studies in relation to four grand challenges:

- ✓ Stakeholder values and ecosystems
- ✓ Technological and infrastructural issues
- ✓ Legal and ethical issues, and
- ✓ Institutional and policy issues.

In the paragraphs that follow, we outline the main findings of each of these horizontal analyses.

The project mapped the **values and motivations** in the development of open access through the elaboration of a functional taxonomy of the stakeholders forming the open access

ecosystem. We identified five basic functions: funders and initiators; creators; disseminators; curators; and users, with stakeholders interacting both within and across functions. This community of stakeholders shares multiple and occasionally overlapping functions and an overarching consensus on the benefits of open access to research data. The benefits relate to the increase in productivity and quality of scientific work, the economic and social values obtained, while there is a clear sense of open access to research data as a general public good. The benefits are therefore seen in the value-context of science as a great value to society, with society benefiting through an on-going dialogue in which knowledge emerges through science as a cumulative process, and the motivations deriving from the above value. Despite this consensus, RECODE showed that the road towards open research data is not perceived in the same way by stakeholders, as a result of 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 access.

Turning to **infrastructure and technology** the project identified five key barriers to successfully implementing open access to research data in Europe: heterogeneity and standardization; accessibility and discoverability; preservation and curation; quality and assessability; and security. Drawing from the RECODE taxonomy, data centres, publishers and libraries are those most affected by the infrastructure and technological issues as a result of their interaction with the infrastructure. Funders' concerns on the other hand focus on sustainability and other policy aspects, such as licensing and certification. Studying the above challenges it became evident that technical and infrastructure solutions should address in a comprehensive way data harmonization, discovery and access, preservation, technological obsolescence, documentation and metadata, quality and relevance indicators and security issues. This should come as a response to the fragmented and often narrow purpose character of current technical solutions for data management and preservation. In applying these solutions, it is important to have in mind the different disciplinary communities and their needs, while also acknowledging and accommodating data variety. In addressing accessibility and sustainability RECODE proposed enforcing the use of persistent digital identifiers, while the promotion of a data management culture and appropriate infrastructure for data curation and preservation was also considered key. In relation to quality, the project highlighted the need to support different stakeholders and communities to develop and deploy complete and accurate metadata (possibly through automated tools). In relation to security, privacy and other ethical issues it highlighted the importance of distinguishing among different levels of open access.



RECODE analyzed a broad range of **legal and ethical challenges**: the former included intellectual property (copyright, trade secret, database right and licensing) and data protection issues (personal data, data minimization, data retention, pseudonymisation, research use exception, consent or alternative legitimate basis, fair processing, right to erasure), while the latter focused on unintended secondary uses and misappropriation, dual use, violations of privacy and confidentiality, unequal distribution of research results, commercialization, restriction of scientific freedom. In examining these challenges, RECODE explored a number of solutions of cross and inter-disciplinary character. As a way of overcoming intellectual property issues, the analysis indicated that the use of licensing models like creative commons and government open licenses should be explored as good practice, but also cautioned on the barriers that might arise from their use. In terms of access management, RECODE supports the elaboration of integrated solutions leading to the homogenization of data management and access practices, while acknowledging that in the absence of universal processes, each

discipline/ institution/ organization/project will have to develop its own tailored processes. A further solution explored was that of editorial/ethical review committees as a way of ensuring ethical data practice and legal compliance, yet once again acknowledging the need to develop alternative practices to adequately attend to disciplinary or organizational specificity. In addressing legal and ethical challenges the report argued that the use of technical, disciplinary or institutional solutions should be encouraged and developed to adequately address these issues.

RECODE also examined four key challenges faced by **institutions**, such as archives, libraries, universities, data centres, and funding bodies. These challenges are: financial support; quality and trustworthiness of data; training of researchers and other relevant stakeholders; and creating awareness about the opportunities and limitations of open access to research data. We found that, although increasingly more institutions are addressing these challenges, there remain a number of barriers to overcome. In relation to funding, for instance, there is lack of understanding of the costs involved in the curation and preservation of the growing volume of diverse research data, while, the long-term funding of data is a further pressing concern. With regard to maintaining the quality of data, the focus of institutions has been primarily on technical quality. However, in order to build trust in the open access infrastructure, institutions also need to encourage and support the development of strategies and tools to ensure scientific quality. Our analysis of institutional challenges has highlighted the need to address the skills gap in libraries and data centres, the better integration of data management and curation skills in post-graduate and educational curricula and the advancement of awareness-raising activities. A key observation is that open access to research data is still very much in the early stages of development, and institutions put on hold solving the harder problems, such as funding long-term preservation of data, evaluating the scientific quality of data or engaging with research communities that do not have a strong digital data sharing tradition. In developing and implementing policies institutions will have to collaboratively take on these problems in order to make open access to research data possible for the diverse scientific and scholarly disciplines.



On the basis of this analysis, RECODE has elaborated ten overarching recommendations. These are intended to be broad, in order to encourage the development of consensus-building and clarify relationships within the open access ecosystem where possible, as well as to enable attention to disciplinary, stakeholder or organizational specificity. 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 with more specific recommendations for each stakeholder category. 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 while developing or implementing policies or programmes.

The RECODE ten overarching recommendations are the following:

1. Develop aligned and comprehensive policies for open access to research data
2. Ensure appropriate funding for open access to research data
3. Develop policies and initiatives that offer researchers rewards for providing open access to high quality data
4. Identify key stakeholders and relevant networks and foster collaborative work for a sustainable ecosystem for open access to research data

5. Plan for the long-term, sustainable curation and preservation of open access data
6. Develop comprehensive and collaborative technical and infrastructure solutions that afford open access to and long-term preservation of high-quality research data
7. Develop technical and scientific quality standards for research data
8. Require the use of harmonized open licensing frameworks
9. Systematically address legal and ethical issues arising from open access to research data
10. Support the transition to open research data through curriculum-development and training

For each of these overarching recommendations, different stakeholders have specific roles to play in achieving the vision. In the sections below, we provide specific recommendations that outline how each stakeholder can contribute to these overarching recommendations, based on the good practice we uncovered in each of the challenge areas and in each of the case studies.

**Research Funders:**

1. Develop explicit policies for open access to research data with clear roles and responsibilities
2. Adopt a comprehensive approach in funding the implementation of open access to and preservation of research data
3. Reinforce the significance of the Data Management Plan (DMP) to embed and promote data management as a distinct activity within the research process
4. Raise awareness and promote open research data in view of leading an open science paradigm
5. Foster collaboration with relevant stakeholders and networks



**Research Institutions:**

1. Develop an explicit institutional research data strategy with open access as the default position
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
3. Include open access to high quality research data as a formal criterion for career progression
4. Develop educational and training programmes for researchers and staff to improve data management skills and to enhance data-intensive research
5. Raise awareness about the benefits of open access to research data and provide rewards
6. Support the research community through the provision of legal and ethical advisory services

**Data Managers:**

1. Assess their position within the open access ecosystem in view of developing collaborative infrastructures and services
2. Develop sustainable business models to ensure long-term service provision
3. Establish mechanisms for data quality that ensure re-use and long-term preservation through collaborative work

4. Acquire certification/accreditation to guarantee high quality services in the long term
5. Support data management through the development of training programmes for researchers and librarians/ technical staff

**Publishers:**

1. Gradually develop mandatory policies for open access to research data supporting publications
2. Collaborate with certified repositories and data centers to streamline data submission
3. Support data as a first-class scholarly output through the establishment of peer-review processes
4. Develop policies requiring citations for research data
5. Establish licensing policies that encourage the use of TDM

# 1. INTRODUCTION

## 1.1 THE RECODE PROJECT AND THE RECODE RECOMMENDATIONS

RECODE (http://recodeproject.eu) 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.



**The policy recommendations** that form the core of this deliverable 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:

- a summary of the RECODE project approach and 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

A short version of this report is available at the RECODE project website (http://recodeproject.eu), along with other reports produced in the framework of the project.

The stakeholder specific recommendations (in an earlier version), were discussed extensively at a workshop that took place in Amsterdam, on September 25th, 2014, in the frame of an RDA plenary meeting. Participants representing all four stakeholder groups to whom the recommendations are directed to, took part in the workshop. The workshop largely validated the results of the project and the direction taken with the recommendations, and participants

offered valuable input. For their participation and suggestions for the recommendations we thank all of the workshop participants.

## 1.2 OPEN ACCESS TO RESEARCH DATA

The discourse on open access to research data is aligned with the notion of rigorous science and the societal and economic benefits obtained. According to the European Commission "open access to scientific research data enhances data quality, reduces the need for duplication of research, speeds up scientific progress and helps to combat scientific fraud".[1] The Royal Society Report sees open inquiry as being at the heart of the scientific enterprise permitting "to identify errors, to support, reject or refine theories and to reuse data for further understanding and knowledge".[2] In such context open data is part of open science with the latter being defined as 'open data (available, intelligible, assessable and usable data) combined with open access to scientific publications and effective communication of their contents'.[3]



The development of open access has also been triggered by technological developments that give rise to new ways of making use of data and new opportunities for communication and collaboration among scientists.[4] The increasing use of computational technologies that allow massive datasets to be analyzed and shared has led to the transition in what has been termed the "fourth paradigm of science" based on data-intensive computing[5] or "data-led science".[6] More specifically, data-led science opens up new sources of knowledge through the development of digital means for producing, storing and manipulating data. This means that informatics are not only used as tools for supporting traditional ways of scientific research in a particular discipline, but more importantly have the potential to change fundamentally the development of a discipline.[7]

The interest in open data and the benefits derived from open access are nonetheless not new; on the contrary open access has been promoted over the past decades through a number of initiatives. Key milestones include the OECD Declaration (2004) on access to research data from public funding[8] on the basis of a number of objectives and principles,[9] followed by the publication of OECD's principles and guidelines a few years later.[10] At EU level the most

[1]European Commission, Recommendation on access to and preservation of scientific information, C(2012) 4890 final, Brussels, 17 July 2012. http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[2]The Royal Society, *Science as an open Enterprise*, London, 2012. p.7. https://royalsociety.org/~/media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf

[3]Ibid. p. 16

[4] For an overview of the key issues in the development of open access to research data, see Wessels, Bridgette, Rachel L. Finn, Peter Linde, Paolo Mazzetti, Stefano Nativi, Susan Reilly, Rod Smallwood, Mark J. Taylor, Victoria Tsoukala, Kush Wadhwa and Sally Wyatt, "Issues in the development of open access to research data", *Prometheus*, Vol. 32, No 1, 2014, pp. 49-66. http://www.tandfonline.com/doi/pdf/10.1080/08109028.2014.956505

[5]Hey, Tony, Stewart Tansley and Kristin Tolle, *The Fourth Paradigm*, Microsoft Research, Redmond, Washington, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf

[6]The Royal Society, op. cit., 2012, p. 7.

[7]The Royal Society, op. cit., 2012, p. 31.

[8]Organization for Economic Cooperation and Development, OECD Declaration on Access to Research Data from public funding, 30 January 2004, C(2004)31/REV1. http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157 .

[9]These are the following: openness, transparency, legal conformity, formal responsibility, professionalism, protection of intellectual property, interoperability, quality and security, efficiency and accountability.

[10]OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD, Paris, 2007.http://www.oecd.org/sti/sci-tech/38500813.pdf .

important initiatives are the Commission Recommendation 'on access to and preservation of scientific information' and the provisions in Horizon 2020. The Recommendation encourages member states to define clear policies accompanied by concrete objectives and indicators to measure progress, financial planning and implementation plans, including the allocation of responsibilities. Horizon 2020 includes a pilot action on open access to research data. The pilot, that covers for the 2014/15 period seven thematic areas[11] and corresponds to about € 3 billion or 20% of the overall Horizon 2020 budget for 2014 and 2015 aims 'to improve and maximize access to and re-use of research data generated by projects'.[12] Open access to research data has also been promoted through an increasing number of reports and roadmaps some of which have been produced in the framework of projects. Examples include the reports produced from projects like Opportunities for Data Exchange (ODE), MedOANet, APARSEN, or the League of European Research Universities (LERU) Roadmap for Research Data.

Nonetheless, progress on open access is rather slow highlighting that the transition to open access is neither easy to achieve nor cost free: it requires investments in infrastructure and technology, and more importantly a change in research culture which takes time to take effect. In relation to the latter, this means that significant variation is expected between disciplines as the development of open access is already common practice in some of them, while important obstacles are still observed in others.[13]



The difficulty in promoting open access is observed even in relation to defining basic notions, i.e. open access and research data. According to the Berlin Declaration open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia materials.[14] In its Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020 the European Commission defines open access as "the practice of providing on-line access to scientific information that is free of charge to the end users and that is re-usable"[15]. In the context of research and innovation, "scientific information" can refer to i) peer-reviewed scientific research articles (published in scholarly journals) or ii) research data (data underlying publications, curated data and/or raw data)".[16] The Royal Society Report defines open data as those that meet the criteria of intelligent openness; i.e., data that are accessible, useable, assessable and intelligible.[17] The OECD defines openness as "access on

[11]These seven thematic areas are the following: 1. Future and Emerging Technologies, 2. Research Infrastructures- part e-infrastructures, 3. Leadership in enabling and industrial technologies-Information and Communication Technologies, 4. Societal Challenges: Secure, Clean and Efficient Energy- part Smart cities and communities, 5. Climate Action, Environment, Resource Efficiency and Raw Materials, 6. Societal Challenge: Europe in a changing world-inclusive, innovative and reflective societies, 7. Science for and with Society.

[12]European Commission, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 1.0, 11 December 2013. p. 8 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf .

[13] Archambault, E., Amyot, D., Deschamps, P., Aurore, N., Rebout, L. & Roberge, G.: Science Metrix Report: Proportion of Open Access Peer-Reviewed Papers at the European and World Levels—2004-2011, August 2013. http://www.science-metrix.com/pdf/SM_EC_OA_Availability_2004-2011.pdf

[14]Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf.

[15] European Commission, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, version 1.0, 11 December 2013 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

[16]Ibid.

[17]The Royal Society, op. cit., 2012.

equal terms for the international community at the lowest possible cost, preferably at no more than the marginal cost of dissemination".[18]

Defining research data is equally difficult, since "any material used as foundation for research can be classified as research data".[19] The OECD uses a wide definition that includes any kind of resource useful to researchers,[20] while the European Commission defines as research data, that which "may be numerical/quantitative, descriptive/qualitative or visual, raw or analyzed, experimental or observational".[21] Definitions also vary with some contributions defining research data as potentially all data –including public sector information- and some limiting it to the product of research.[22] A further distinction is that between open research data and open data, with the latter mainly used in the context of open government initiatives.[23]

The above analysis indicates that notwithstanding the variability and differences in defining open access and research data, an increasing push is observed in developing and promoting further open access at both national and supranational level.



## 1.3 THE RECODE APPROACH

Despite the general consensus amongst policy makers on the benefits of open access for science, industry and civil society, there are still important barriers that need to be overcome. The RECODE project 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 forms of data collections. Against this background, the report provides policy recommendations on open access to research data targeted at key stakeholders in promoting open access: research funders; data managers; research institutions; and publishers. In other words, the recommendations are targeted to those stakeholders within the open access ecosystem who have been identified in the framework of the project as instrumental in bringing about change.[24]These recommendations will assist the above stakeholders in furthering the goals of open access to research data in each of their organizations and networks. Recommendations include both overarching and stakeholder-specific ones as suggestions to address and attend to these two issues by building on or learning from existing experience. In doing so, RECODE recognises that in some situations it is appropriate to build consensus and transfer good practice across disciplines and stakeholder

[18]OECD, op. cit., 2007, p. 15.

[19]Sveinsdottir, Thordis, Bridgette Wessels, Rod Smallwood, Peter Linde, Vasso Kalaitzi and Victoria Tsoukala, *Stakeholder Values and Ecosystems*, D1.1 RECODE Project, 30 September 2013. http://RECODEproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf.

[20]Organisation for Economic Cooperation and Development, op. cit., 2007.

[21] European Commission, A Reinforced European Research Area Partnership for Excellence and Growth, COM(2012) 392 final, Brussels, 17.07.2012, file://filesrv/userdocs2/mangelaki/My%20Documents/Downloads/AReinforcedEuropeanResearchAreaPartnershipforExcellenceandGrowth.pdf .

[22]Bigali, Lorenzo, Thordis Sveinsdottir, Bridgette Wessels and Rod Smallwood, Peter Linde, Jeroen Sondervan, *Infrastructure and technology challenges*, D2.1 RECODE Project, 31 March 2014. http://RECODEproject.eu/wp-content/uploads/2014/04/D2.1-Infrastructure-and-technology-challenges.pdf.

[23]Heinz Pampel and Sunje Dallmeier-Tiessen, Open Research Data: From Vision to Practice, Bartling Sonke and Sascha Friesike (eds.) Opening Science, The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing, 2014.

[24]The mobilization of key stakeholders is a central issue in bringing about change. It has been addressed in RECODE WP6, cf. Linde, Peter, Bridgette Wessels, Rod Smallwood, Merel Noorman, Sally Wyatt, Jeroen Sondervan, Feasibility of using existing open access networks to support the harmonization of open access: A report examining the feasibility of using existing open access networks to support the harmonization of open access policies across Europe, D6.1 RECODE Project, forthcoming.

groups, while in others it is appropriate to enable and support specific groups to maintain their particularity in relation to disseminating, preserving and re-using research data. In addition, for each stakeholder group good practice examples are provided that can serve as "models" for providing open access to research data. Finally, the recommendations and good practice examples will reduce "costs" associated with providing open access, as stakeholders do not need to develop their own expertise, but can use this information as a foundation to develop their own policies, support actions and initiatives. In such a context researchers, as the producers of data and consequently the origin point of the data lifecycle, emerge as an essential component of open access processes whose needs, concerns and interests must be considered in order to work towards a strong open data ecosystem.

As such, the development of open access to research data needs to be:

- ✓ Characterised by a partnership approach involving the key stakeholders, researchers, and institutions
- ✓ Supported by an integrated institutional and technological data infrastructure
- ✓ Guided by ethical and regulatory frameworks
- ✓ Informed by research practices and processes in different fields

Developing open access to research data in a way that is informed by awareness of differences in research practices within and between disciplines and characterised by a partnership approach among key stakeholders helps to ensure engagement from the wide range of research communities and to better embed it within research practice and process.



The policy recommendations presented here are informed by these overall findings as well as a series of horizontal analyses of the RECODE case studies in relation to four grand challenges:

- ✓ Stakeholder values and ecosystems
- ✓ Technological and infrastructural challenges
- ✓ Legal and ethical challenges, and
- ✓ Institutional and policy challenges.

In formulating the recommendations, RECODE incorporated the results that the project produced on the above four areas of work and examined from the perspective of the stakeholders to whom the recommendations are addressed. Furthermore, it conducted a review of scholarly literature, policy documents, and reports, significant work of other EC-funded projects (e.g. ODE, APARSEN, PARSE.Insight) and relevant documents to provide an overview of the current policies, practices and challenges for these stakeholders both in the EU and abroad to making open research data open, identify issues of importance and concern and present institutional solutions to these issues.

The present section provides a short presentation of the work carried out by the project, which in turn informs the recommendations, arranged by sequence of work.

**Stakeholder Values and Ecosystems**[25]

The first grand challenge that RECODE analyzed was the stakeholder values and ecosystems. Identifying the stakeholders in research and data ecosystems is of particular importance as it supports the development of a coherent approach to open data while it also ensures that the

[25]This section draws on Sveinsdottir et al., op. cit., 2013, p. 21.

policy recommendations are informed by practice. Their number is quite large and their nature diverse, including (among others) national governments, the industry, the public, mass media, publishers, scholarly and professional societies. This complex ecosystem has been studied through the development of a functional taxonomy, in essence re-structuring the stakeholder list around a number of broad functions. The functions (or entities) identified are: 1) funding and initiating, 2) creating, 3) disseminating, 4) curating, and 5) using, all of which are interconnected through flows. The open access ecosystem is thus formed of the above community of stakeholders with multiple functions, yet sharing the same overarching values and motivations with regard to open access.

According to the proposed taxonomy, each stakeholder can have multiple functions (primary and secondary), yet it is the primary function that defines a stakeholder's position within the open access ecosystem. A primary function performer is a performer with an essential importance to the function, while secondary stakeholders are performers not essential to the function. Whereas each stakeholder can only have one primary function, the existence of multiple secondary functions (along with the primary one) results in an over-lapping of functions. This over-lapping contributes to the further blurring of responsibilities, and renders the development of open access policies a challenging task as it involves numerous stakeholders with multiple and inter-related responsibilities.

The added-value of the RECODE proposed taxonomy is that it allows us to understand the complexity of the ecosystem and to gain an overview of the values and motivations of open access to data across the different stakeholders (including researchers). In addition, it permits us to understand that the issues addressed throughout the project are not always experienced or interpreted in a uniform way from the different stakeholder groups as a result of their different functions in the ecosystem. Thereby, the ambition to make more research data openly available is not straightforward, requiring considerable effort and involving different phases like ingestion, storing or providing access.



On the basis of the RECODE functional taxonomy we have identified four stakeholder groups as key actors in promoting and bringing about change in open access policy implementation for research data: 1) funders, 2) data managers, 3) research institutions, and 4) publishers. While the project acknowledges the existence of a large number of stakeholders in the open access ecosystem, we focus on the above-mentioned four categories given their central role in the ecosystem and their capacity for bringing about change, and thus the RECODE recommendations are targeted to these stakeholder categories.

In relation to the values and motivations, there is an overarching consensus among the community of stakeholders on the value of making data open.[26] The benefits of open access to research data relate to the increase in productivity and quality of scientific work, the economic and social benefits obtained, while there is a clear sense of open access to research data as a general public good. The benefits are therefore seen in the value-context of science as a great value to society, with society benefiting through an on-going dialogue in which knowledge emerges through science as a cumulative process, and the motivations deriving from the above values. Nonetheless, despite this overarching consensus, stakeholders within each functional area are also aware of the practical issues entailed in developing open access (namely the key

[26] More extensive on the benefits of data sharing, the report just released by the RDA: The Data Harvest. How Sharing Research Data Can Yield Knowledge, Jobs and Growth. A special report by RDA Europe, 2014, https://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

challenges studied in the framework of the RECODE project). Values and motivations are also understood differently by researchers, and from within specific disciplinary and interdisciplinary perspectives as different disciplines vary in their degree of openness, on how they share within their community and their overall attitude towards open access.

From a researcher's perspective five concerns have been identified as potential barriers to developing open access to data: competition for prestige and funding, the amount of work involved in making data meaningful in open access, the limited value that publishing data receives as a valuable scientific activity, the concerns in relation to making sensitive data open, and the need to provide access to the context in which data is analysed and collected. This means that developments in open access have to be sensitive to the specific processes of scientific practice to ensure that the research rigor is maintained while facilitating open access.

**Infrastructure and Technology Challenges[27]**

The second grand challenge focused on infrastructure and technology. 'Infrastructure' includes technological assets (hardware and software), human resources, procedures for management, training and support to its continuous operation and evolution. The RECODE report on infrastructure and technology concluded that technological challenges are not viewed as a high concern in implementing open access to research data when compared to financial, cultural and legal ones.

The main infrastructure and technology challenges identified by the project were grouped in five broad categories: heterogeneity and interoperability; accessibility and discoverability; preservation and curation; quality and assessability; security. Heterogeneity and interoperability cover issues that arise because of different ways of formatting, storing, operating and standardizing data, thus rendering seamless open access to research data a complex technological undertaking. The importance of interoperability lies in the fact that it allows data exchange between researchers, institutions, organizations, countries etc., while further benefits are derived by producing deeper and better-integrated understanding. Closely related to interoperability issues is the sustainability of research infrastructures; as many data centers rely on short-term funding, there is a danger of datasets getting lost in the event of not being able to secure follow-on funding.



Issues of accessibility and discoverability highlight the need for metadata standards and standard data formatting. Measures addressing the problems related to data discovery and access include digital object identifiers (DOI) and persistent identifiers to data publishers, datasets, the data record itself, data versioning and data citation, data usage index (DUI) and effective data citation mechanisms.

Preservation and curation have already been identified as technical and infrastructural barriers inhibiting the sharing of research data. To address this challenge investment in long-term preservation must be undertaken along with efforts to keep hardware and software up to date. Delegation of responsibility for data storage and accessibility to neutral institutions (national institutional repositories or digital libraries) has also been confirmed by the on-line survey conducted within the framework of the study of technology and infrastructure challenges. Yet, technical solutions for data management and preservation are often fragmented and designed for a narrow purpose, rather than adopting a more centralized approach.

[27]This section draws on Bigagli et al., op. cit., 2014.

Finally, quality, assessability and security issues require (among others) the establishment of processes for ensuring quality standards, development of appropriate education and training material, certification schemas and accreditation processes.

The identification of the above barriers points to the central role repositories, libraries and publishers have in developing appropriate solutions.

**Legal and Ethical Challenges**[28]

The third grand challenge analyzed legal and ethical issues. The project examined in particular both legal issues such as intellectual property rights, including copyright, trade secrets and database rights, privacy and data protection and open access mandates and ethical ones including the unintended secondary use, misappropriation and commercialization of research data, unequal distribution of scientific results and disproportionate impacts on scientific freedom. In particular the project looked into how these different issues impact on a range of stakeholders such as policy-makers, researchers, repository managers, and institutional representatives.

As shown from the RECODE analysis, intellectual property rights, especially in relation to data that has been purchased from commercial organisations or cultural data, can act as a significant barrier to providing open access to research data, as sometimes the data creators may not hold the intellectual property rights to the material they collect and to which they seek to provide access. Similarly, research participants, rather than researchers, institutions, repositories and other stakeholders, have primary control over the use of personal information for research purposes, which can limit the extent to which this data can be made available in open access. Furthermore, these legal regimes often create a complex landscape, with real consequences for researchers, organisations and institutions. Open access mandates from governments and funders may place researchers and institutions in a situation where they are pressured to provide open access to data, despite the fact that intellectual property rights or data protection rights specifically and explicitly limit their ability to do so.



Open access to research data raises several ethical concerns as well. Many echo or exacerbate existing concerns about sharing research data in general. For example various disciplines have formalized principles on ethical research in their codes of ethics urging their researchers to treat data confidentially and to ensure that the benefits and benefits of research are equally distributed. Failing to meet these ethical standards may not only cause harm to research participants, but can also prove detrimental to the scientific enterprise or society. Open access to research data raises concerns about the ability of researchers to adhere to these standards and the disruptive effects it may have on existing infrastructures and practices. Unintended secondary use can damage identities, reputations and relationships between individuals, and may even endanger research subjects or sites as well as the public trust in science or social institutions. The valid concerns described above are not necessarily reasons to avoid providing open access altogether. In some cases, the benefits of providing unrestricted access to data can offset the potential risks.

A further issue examined relates to privacy, as open access and privacy seem difficult to reconcile. Anonymisation of data does not suffice to mitigate the risk for all data sets. As a

[28]This section draws on Rachel Finn and Kush Wadhwa, Mark Taylor and Thordis Sveinsdottir, Merel Noorman and Jeroen Sondervan Legal and ethical issues in open access and data dissemination and preservation, RECODE project, Deliverable D3.1, 2014 http://RECODEproject.eu/wp-content/uploads/2014/05/D3.1-legal-and-ethical-issues-FINAL.pdf .

result of technological advances and the availability of increasingly more digital data sets, anonymisation can be more easily undone, while in some cases it is not even possible because the data content enables identification and resists effective obfuscation. Finally, open access to research data can level the playing field, but there is no guarantee that all stakeholders will benefit equally. It may reinforce or even lead to an unequal distribution of those results. Those who lack the required scientific, technical or cultural capital and resources to make use of data are at a disadvantage, even when the data are formally open to all.

Overall, the analysis concluded that while new solutions should be sought to provide legal and ethical pathways to open access, the current push must accept the existence of some limits and caveats that may be related to intellectual property or data protection and ethical research practice.

**Institutional Challenges[29]**

The final grand challenge addressed focused on institutional issues. 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 have all been identified as key challenges faced by institutions such as archives, libraries, universities, data centres, and research funders.

In relation to financing, while the potential cost-savings have been central to the discourse surrounding open access, the latter is not cost-free, thus placing an important burden on institutions as a result of their ability to secure the necessary funds for open access to research data. Funds are needed for the various phases in the data life cycle project including preparation, ingestion, sharing and archiving. Such costs are not only related to projects generating large data sets but can be quite heavy even for smaller individual projects. Technological developments may place an additional burden as the introduction of new practices may necessitate the need to secure further funds. At the same time, it is not always clear as to who shall bear the related costs, while institutions are being under increasing pressure to define the costs related to data management. Measures to address financing issues include the creation of economies of scale through multi-institutional collaborations, cost modeling and exchange of information on costs and the development of funding models that take into account the long-term curation of research data.



Ensuring the quality of data is a further issue of importance as researchers need to have some level of confidence in the accuracy and soundness of open research data. High quality of research data is an integral component for the ability to reuse data. In many disciplines, formal and informal mechanisms are in place to check the quality of research data produced, but open access to research data often requires additional mechanisms, for instance, to ensure that data are re-usable and interpretable. Several stakeholders are involved in these processes, including, beyond the researchers, data repositories and data centers, research consortia, and publishers. While institutions have developed various strategies such as peer- review procedures, citation records, standard metadata, transparent review and publishing practices issues still remain. Solutions for addressing this challenge include the further development of research cultures in which data is an integral part of the evaluation system, the development by journals of standards, methods and criteria for reviewing data effectively, along with the development of

[29]This section draws on Noorman Merel, Vasso Kalaitzi, Marina Angelaki, Victoria Tsoukala, Peter Linde, Thordis Sveinsdottir, Lada Price and Bridgette Wessels, Institutional Barriers and Good Practice Solutions, RECODE Project, Deliverable D4.1, 2014. http://recodeproject.eu/wp-content/uploads/2014/09/RECODE-D4.1-Institutional-barriers-FINAL.pdf .

new publishing products (namely data journals and data articles) that contribute in enhancing the quality standards of research data or finally the use of altmetrics.

Institutions are also expected to play a key role in providing training to both 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.

### 1.4 OVERARCHING RECOMMENDATIONS

On the basis of this analysis, RECODE has elaborated the following **ten** overarching recommendations. They are intended to be broad, in order to encourage the development of consensus-building and clarify relationships within the open access ecosystem where possible, as well as to enable an attention to disciplinary, stakeholder or organizational specificity. 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 with 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 related policies or programmes. A short version of the overarching and stakeholder-specific recommendations are presented in the booklet *Policy Recommendations for Open Access to Research Data*, which is available on the project website (www.recodeproject.eu).



The RECODE overarching recommendations are the following:

1. ***Develop aligned and comprehensive policies for open access to research data***
   Policies should be consistent with national priorities and aligned with the European framework for open access to research data (2012 EC Recommendation on open access), while also complementing that for open government data. No mandate should exist without providing the appropriate infrastructure and services that allow researchers to comply with funder requirements.

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. Provision should be made for funding the establishment 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 providing open access to high quality data***
   Funder and institutional policies that provide researchers with a suitable reward structure for open access to research data are central in the transition to an open access environment. Funders are encouraged to reward researchers through the formal recognition of data sharing in their funding decisions. Research institutions are also encouraged to include open access to high quality research data as a formal criterion for career progression.

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, collaborative work is essential as it allows the gradual development of a coherent view among stakeholders, an agreement on their roles and responsibilities, the allocation of resources and the alignment of stakeholders' policies, while avoiding the duplication of effort and loss of resources.

5. ***Plan for the long-term, sustainable curation and preservation of open access 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***
   Such solutions should 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***
   In promoting the importance of complete and accurate metadata institutions should make use of a variety of strategies, including automating part of the ingestion and quality control process, and manual quality assurance. In addition to evaluating the technical quality of research data research communities should collaborate in developing appropriate strategies for the evaluation of the scientific quality of data.



8. ***Require the use of harmonized open licensing frameworks***
   Open licenses, like creative commons, provide a useful way on how information 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 for a, 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 establishing curricula that contribute towards the development of data science and information management as distinct and legitimate career paths.

# 2. THE RECODE STAKEHOLDER RECOMMENDATIONS

## 2.1 RESEARCH FUNDERS

### *Context*

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 European and national governments, such as the relevant ministries that develop national policies and strategies and distribute base research funding, individual public funders that distribute competitive funding on a national basis, non-profit institutions and private funders. Understandably, there is 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. This results in different measures and strategies, ranging from national laws to recommendations (EC) and funder policies of varying effectiveness.

While 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, [30] funder policies for open access to research data remain limited, especially when compared to those applicable to peer-reviewed publications. [31]At the EU level, the most important funder is undoubtedly the European Commission (EC), representing an important source of competitive funding for some member states. On the basis of the size of allocated funds, the EU 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"[32] 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 affects 20% of the overall Horizon 2020 budget in 2014 and 2015, or approximately 3 billion euros, and is implemented in seven areas with the aim to improve and maximize access to and re-use of research data generated by the projects.[33]



[30]Sveinsdottir et al., op. cit., 2013 p.13.
[31] Caruso, J., Aurore N. and Archambault, E. *Open Access Strategies in the European Research Area,* Science Metrix, August 2013. http://www.science-metrix.com/pdf/SM_EC_OA_Policies.pdf
[32]European Commission, op. cit., 2012.
[33]Details in the European Commission's Pilot Guidelines http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

**Article 29.3 of Model Grant Agreement**

**Projects participating in the Pilot on Open Access to Research Data in Horizon 2020**

Regarding the digital research data generated in the action ('data'), the beneficiaries must:
(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
(ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan';
(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).
As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardized by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.

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



At member state level, UK research funders are global pace-setters in policy development for research data and in comprehensively developing the relevant services. RCUK, the strategic partnership of the UK's seven Research Councils has adopted open access policies since 2005.

**RCUK Common Principles on Data Policy**

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice
- To enable research data to be discoverable and re-used, sufficient metadata should be recorded and made openly available
- Research organization policies and practices should ensure that legal, ethical and commercial constraints on release of research data are considered at all stages in the research process
- Those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research
- All users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data

Source: http://www.rcuk.ac.uk/research/datapolicy/

Some of the UK research councils had already adopted open access policies for research data before 2011, but following the publication of the common policies all seven of them have now adopted open access policies to research data that align to the aforementioned principles.[34] These policies have had a significant impact on universities in the UK, as many of them have rushed to develop the organizational and technical infrastructure and services that will allow their researchers to meet funder expectations.[35]For example, EPSRC developed a policy requiring UK institutions to develop mandatory open data policies and the relevant infrastructures for open access to the research data produced in the framework of EPSRC funded projects within a specific timeframe. As the EPSRC provides the largest amount of funding for research to UK institutions, the latter immediately sought to implement the requirements of the funder. This policy has demonstrably been very impactful in institutional change and researchers' cultures by embedding research management and open access to research data in the research process.[36]

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.[37]There are, however, exceptions and some funders are gradually moving toward this direction. In Ireland, the national principles for open access, agreed on the basis of a consensus-building process, provision open access to research data as a default when this is possible.[38] The Austrian Science Fund, the main research funder of the country, requires open access to research data whenever legally or ethically possible in appropriate repositories[39] while the DFG (the German Research Foundation) encourages deposit of research data in appropriate repositories.[40] In the fall of 2014 the Research Council of Norway issued its policy on open access to research data for research partially or wholly funded with public money. The policy is not mandatory and comes in the form of principles and guidelines which apply to all data generated by projects funded by the Research Council, with some exceptions (security concerns, personal data etc.).[41]



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.[42] This was a swift and all-encompassing move on the side of the US government to expedite open

[34] Overview and links are presented at the relevant page of the Digital Curation Centre, http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies. An analysis of the policies of the RCUK on the sharing of research data see Sarah Jones, *Developments in Research Funder Data Policy*, International Journal of Digital Curation (2012), 7(1), 114–125, http://dx.doi.org/10.2218/ijdc.v7i1.219 .

[35]This is clear in the analysis presented by the Digital Curation Centre and in Jones, op. cit., 2012.

[36] EPSRC, "EPSRC policy framework on research data" http://www.epsrc.ac.uk/about/standards/researchdata/expectations/ .

[37]Caruso et al., op. cit., 2013.The lack of policies addressing open access to research data is further evidence at a glance in Sherpa Juliet, the registry for funders' open access policies, http://www.sherpa.ac.uk/juliet.

[38]http://www.iua.ie/governement-launches-national-open-access-statement-23-oct-2012/ .

[39]FWF, "Open access policy for FWF-funded projects".https://www.fwf.ac.at/en/research-funding/open-access-policy/ .

[40]http://www.sherpa.ac.uk/juliet/index.php?fPersistentID=5 .

[41]http://www.forskningsradet.no/en/Newsarticle/Research_data_must_be_shared/1254000848864/p1177315753918 (newspiece with pdf of English translation of policy).

[42]The White House, "Expanding Public Access to the Results of Federally Funded Research".http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research .

access to research data funded with public money, which required the development of harmonized policies among different public funders in the US. Universities rushed to comply with the upcoming funders' requirements in developing a common infrastructure to support this policy, the SHARE infrastructure.[43] Nonetheless, this came after a few years of work on Interagency Coordination (Interagency Working Group on Digital Data, 2009), and after two of the major US research funding agencies, namely the NIH and the NSF had developed policies for open access to research data (data sharing policies). The NIH has the oldest existing policy, since 2003,[44] while the 2010 NSF policy[45] resulted in the immediate mobilization of research institutions (especially libraries) to assist their researchers in complying with the requirements for open access to research data.[46]

In Canada a number of research funders require open access to research data they fund. For example, Genome Canada has a strict policy for open access to research data produced by the research it funds since 2008.[47] In Australia, the Australian Research Council "encourages researchers to consider the benefits of depositing their data and any publications arising from a research project in appropriate subject and/or institutional repository wherever such a repository is available". Research outputs that have been or will be deposited should be indicated in the final report.[48]



The aforementioned practices and developments 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 and empowering the international nature of research to flourish, policy alignment is very important, and therefore awareness of international developments and alignment to them (through e.g. RDA, Science Europe etc) is crucial, as well as solid knowledge of own country status with reference of infrastructure, support services and scholarly practices.

With respect to the policy requirements and content, a survey of the major funder policies in Europe and the USA shows the existence of important common requirements that are useful to highlight, in particular for funders interested in developing policies. These have been translated into practical checklists at the end of this document in view of assisting this process. In terms of the form of the policy, 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. They require deposit of research data supporting publications and other important research data in certified repositories within a specific timeframe (either simultaneously with publications or by the end of the project). They require researchers to describe these and other provisions (e.g. evaluation of their data; long term preservation provisions) in mandatory

[43]SHARED Access Research Ecosystem.http://www.arl.org/storage/documents/publications/share-proposal-07june13.pdf .

[44]http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html .

[45]National Science Research Foundation, "Dissemination and Sharing of Research Results".http://www.nsf.gov/bfa/dias/policy/dmp.jsp .

[46] Dietrich, Dianne, Trisha Adamus, Alison Miner and Gail Steinhart, "Demystifying the research data management requirements of research funders", Issues in Science and Technology Librarianship, Summer 2012, DOI:10.5062/F44M92G2 .

http://www.istl.org/12-summer/refereed1.html .

[47]Genome Canada. "Data Release and Resource Sharing".http://www.genomecanada.ca/medias/PDF/EN/DataReleaseandResourceSharingPolicy.pdf .

[48]Australian Research Council, "Discovery Projects. Funding rules for funding commencing in 2008".http://www.arc.gov.au/pdf/DP08_FundingRules.pdf .

DMPs, which are submitted with the grant proposals and evaluated.[49] 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 the policies (e.g. DOI, specific metadata standards etc.)[50]as well as provisions on appropriate licensing. The EPSRC for example stipulates that metadata should be enough to allow others to understand what research data exists, why, when and how it was generated and how to access it, should include a DOI when accompanied by a digital object, and explain why data is not accessible in the event of obstructed access.[51] 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.

**The ESRC Research Data Policy**

The ESRC Research Data Policy, published in 2010 and effective as of 2011, is a mandatory policy which requires grantees to deposit their research data and make them openly available in a timely and responsible way specifically through the UK Data Service (UKDS). UKDS has been designated to curate data in Social Research and receives direct support from the ESRC. The ESRC policy, which is characterized as very detailed, assigns responsibilities to all parties involved, provides full guidance to researchers on their obligations and addresses strategic issues such as IPR, copyright and confidentiality, security, and ethical considerations. The relationship between the ESRC and the UKDS is particularly close, with the latter providing training to grantees on data management and how to prepare their data, as well as evaluating research data for ingestion. The ESRC requires that DMPs undergo peer review and thus also provides extensive guidelines for the reviewers. ESRC addresses the issues of the quality of the research data to be curated and opened up, through the DMP, and emphasizes in its policy that 'research data must be accompanied by high-quality metadata in order to provide secondary users with the important additional information, for example, the origin, circumstances, processing/analysis and/or the researchers' management of the data'. The ESRC declares that it monitors its policy and reserves the right to withhold payments for non-compliance. Source: http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx



On the basis of their role in driving forward research, funders assume a new and increased responsibility to ensure that funding is available for infrastructure and services for access to and long-term preservation of data and for boosting skills and innovative research. Funders, especially public funders, are the main supporters of the infrastructure for open access and long term preservation of the scientific output, including research data. Repositories and data centers have emerged with the financial support of major public funders in various countries in the past

[49]Some funders, such as the Wellcome Trust, require DMPs only for research data that researchers deem as valuable, thus addressing the issue of data evaluation and data selection in contexts of large volumes of data. A report by the Expert Advisory Group on Data Access finds that DMPs are often not reviewed or resourced adequately, while their delivery is not routinely monitored or enforced (Expert Advisory Group on Data Access, Establishing incentives and changing cultures to support data access, May 2014. http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/WTP056496.htm .

[50]Cf. blogpost by Cameron Neylon on the ability to monitor policies http://blogs.plos.org/opens/2014/09/03/policy-design-and-implementation-monitoring-for-open-access/ (Policy Design and Implementation Monitoring for Open Access, 3 September 2014).

[51]EPSRC, "EPSRC policy framework on research data" http://www.epsrc.ac.uk/about/standards/researchdata/expectations/

years, either as national efforts (e.g. the UKDS) or as international collaborations (the ATLAS project by CERN). Currently, most funding models are project-based, and more attention needs to be paid to long-term preservation and, effectively, to assessing and enhancing expertise of data centers of impact, where this is necessary.[52] For the latter, funders may choose to offer financial support to data centers and repositories of national scale (such as for example in the UK (UKDS, NERC data centers), Australia (ANDS), the Netherlands (DANS)) and/or by developing international collaborations for data facilities, to develop economies and efficiencies of scale.[53] This observation highlights the importance of collaboration between funders and data managers. The UK provides a good example of how policies are put into practice: research councils have endorsed the use of public funds for data management and sharing, while JISC has made investments in research data management projects and HEFCE has committed significant funds to shared services.[54]

Re-use is also enhanced through the funding of schemes that promote and develop new insights, tools, standards and practices, ultimately enhancing the re-use of data in new and innovative ways, such as for example initiatives developed by the NIH and the NEH for biomedical sciences and the Humanities, respectively.

**NIH and NEH initiatives to support new types of data research**

The NIH initiative on **Big Data to Knowledge (BD2K)** is directed to biomedical scientists. BD2K has the following aims:

- To facilitate broad use of biomedical digital assets by making them discoverable, accessible, and citable.
- To conduct research and develop the methods, software, and tools needed to analyze biomedical Big Data.
- To enhance training in the development and use of methods and tools necessary for biomedical Big Data science.
- To support a data ecosystem that accelerates discovery as part of a digital enterprise

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

The **Digging into the Data Challenge** begun in 2009 by the NEH, and is targeted to the Social Sciences and the Humanities (SSH). It offers funding for new insights, tools and skills in the SSH by analyzing large-scale data. The initiative is now supported by a total of ten funders also from Europe, who pool resources to this end.

Source: http://www.diggingintodata.org/



[52]On this problem cf. Noorman et al., op. cit. 2014, chapter 2, p. 35.
[53]On this approach in funding and examples, cf.Noorman et al., op. cit. 2014, p.29-30.
[54]Jones, op. cit., 2012.

Beyond the actual data management and curation and in view of being able to perform data-intensive research, funders turn their attention more intensely towards enhancing the data skills of researchers and information specialists, enabling new and innovative research, involving the industry both in funding research, as well as in encouraging new and innovative data initiatives. Acknowledging the importance of innovative actions, the EU under its Horizon 2020 programme has published a call expected to fund innovative actions on big data that will assist European companies in turning large data volumes in semantically interoperable data assets and knowledge.[55]

Finally, since the turn to open access policies for research data is quite recent, evidence regarding the effectiveness and impact of such policies is understandably lacking at the moment. This points to the need for developing the mechanisms and reserve funding to monitor and measure the effectiveness of their policies, which will afford them more evidence-based policy decisions in the future. In doing so, collaboration with other stakeholders may be necessary, and in particular publishers and data centers who can develop technical solutions for measuring policy compliance.[56] Such evaluations are already underway in assessing open access policies for research publications, e.g. by the RCUK. [57]

### *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. Policies should accommodate the possibility for closed data when ethical, copyright, confidentiality, security and similar issues are demonstrably of key concern, and they should take into consideration different disciplinary practices. Funders should require the use of open licenses, such as Creative Commons licenses, as well as machine-actionable schemes, in view of enabling the proliferation of reusable research data. Policies should include robust monitoring and compliance mechanisms and non-compliance should incur consequences, such as the withholding of funding. They should be accompanied by mechanisms for evaluating their effectiveness, allowing in turn their readjustment, if deemed necessary. Funders should fund and/or oversee that appropriate infrastructure and processes are in place, which in turn permit the implementation of mandatory policies for open access to research data.

2. ***Adopt a comprehensive approach in funding the implementation of open access to and preservation of research data***
   Funders are encouraged to provide appropriate resources for the following important issues: 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; the development of skills among researchers and information specialists. Funders, especially major public research funders of nation-wide scale, are encouraged to



[55] http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/87-ict-15-2014.html
[56] Cf. Cameron Neylon, op. cit. 2014 on a new tool funded by PLOS that checks licenses for individual articles to assess compliance with relevant funder requirements.
[57] RCUK. "Data required by RCUK to support compliance monitoring". http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/ComplianceMonitoring.pdf .

mobilize complementary funding instruments, where available, for example from structural funds, the European Commission and the private sector to support the implementation of comprehensive policies for open access to research data. In doing so, funders are encouraged to evaluate the services for research data access and preservation in their country along with the needs of the research community.

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

The DMP has emerged as an important tool through which to address data management systematically as early as the planning stage of research projects. Funders should require a DMP for research data produced by projects and provide relevant guidance on the points that a DMP should address. In strengthening the significance of the DMP, and by extension research data management and sharing overall, funders should acknowledge it as a distinct and required activity, starting early in the research planning process and developing throughout it. Funders should allocate relevant resources for the delivery of DMP plans as part of the funded research and should seek the professional expertise of data managers in developing DMPs and the relevant requirements and monitoring processes. It is expected that the widespread requirement of DMP will gradually lead to cultural change among the research community regarding research data.

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

Funders are expected to take a proactive role in changing research cultures towards an open science paradigm also through (soft-type) activities that impact on research practices and research cultures, such as promoting good practices by specific researchers and research groups and/or establishing prizes for good practices in sharing research data, also in view of providing incentives to researchers.



5. ***Foster collaboration with relevant stakeholders and networks***

Collaboration with relevant stakeholders is key in promoting open access. Funders should take the lead in bringing together researchers, research institutions policy makers, data mangers, and publishers, in view of developing aligned policies and sustainable strategies and infrastructures for open access to research data. Intense collaboration will also contribute in defining in a clearer way the roles and responsibilities of each stakeholder, as well as to address researcher perspectives to policies. International collaboration is equally important, through participation in alliances such as the RDA, Science Europe, that ensure alignment of policies and practices at the international level and facilitate further international collaboration.

## 2.2 RESEARCH INSTITUTIONS

### *Context*

Research institutions refer to universities and higher education organizations engaged in primary or secondary research and to publicly and privately funded research institutes/centres that produce research in national or international constellations (for example CERN or the Max-Planck Institute). Research institutions hold a focal role in transitioning to open access practices, as they are the primary loci where researchers carry out and publish their work.

In recent years, research institutions around the world have been heavily involved in promoting the uptake of open access practices by developing policies, infrastructures and services to support them, as shown in the steadily increasing number of institutions adopting relevant policies.[58] Nonetheless, the main focus thus far has been on open access in relation to publications, with policies for open research data being far fewer.[59] Where relevant policies exist, open access to research data is addressed in the framework of research data management policies rather than as a policy specifically addressing open access to research data. Insofar as research institutions perceive research data as a valuable university asset, they are motivated to develop policies to safeguard their intellectual, financial, human and material investment, to help build more research on it, and to contribute to a better understanding of their contribution to the public good.[60] Furthermore, motivation to develop policies and strategies derives from the increasing pressure from research funders who require demonstration that the research data produced with their funding is properly managed and is, in principle, openly accessible.[61] In some cases however, the motivation for developing a sound institutional data strategy derives from the researchers themselves, who acknowledge the significance of research data and the need for better management.[62]



The UK, the USA and Australia are –once more- the countries where the most consistent progress in institutional strategies and policies for research data management is observed. Rapid developments both in the UK and the USA are mostly the result of funder mandatory requirements: RCUK in the UK and NSF and NIH in the USA.[63]One of the earliest institutional data management policies in the UK is that of the University of Edinburgh, which was approved in 2011, and subsequently became a basis on which universities in the UK and elsewhere built their policies. In Australia, while the policy of the main funding agency is not mandatory for research data, Australian universities have made significant progress in instituting relevant policies, strategies and services to address research data management under the influence of the Australian Code for Responsible Conduct of Research, where having an

[58]Cf. the growth of institutional mandates for open access http://roarmap.eprints.org/ .
[59] Aurore Nicol, Julie Caruso, & Éric Archambault, Science Metrix. Open Data Access Policies and Strategies in the European Research Area and Beyond, 2013 p. ii. http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf
[60] Erway, Ricky. 2013. Starting the Conversation: University-wide Research Data Management Policy. Dublin, Ohio: OCLC Research. http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf .
[61]This is explicit in the institutional data management policy documents of universities, especially in the UK, such as Oxford and Edinburgh http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf (Oxford) http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy (Edinburgh).
[62]This is reported, for example, in the case of Oxford University. Cf. J. Wilson, LIBER Case Study: University of Oxford Research Data Management Infrastructure, June 2014 http://libereurope.eu/wp-content/uploads/2014/06/LIBER-Case-Study-UOX.pdf .
[63]Cf. above section on funders.

institutional data management policy is a requirement.[64] The Horizon 2020 policies for open access to research data and the relevant 2012 Recommendation by the European Commission are also expected to have an impact in the development not only of national but also of institutional strategies and policies in EU member states.

An effective data policy is of utmost importance since it sets the pace and the requirements by which the research community within the institution is to abide. Related policies allocate in a clear way responsibilities and tasks to the different actors within the institution, with the 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. The allocation of responsibilities for each stakeholder is clearly important, yet policies should be flexible enough to accommodate for the changes in researchers' needs and keep pace with technological developments. Institutional policies also 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 for the institution's research data, where this is appropriate and legally possible; they require researchers to devise and follow a DMP for their work; they render researchers responsible for the data management within their project; they acknowledge that funder requirements should be respected. Furthermore, they set requirements regarding where to deposit research data (e.g. institutional or international disciplinary repository), and outline broadly the data retention policy/strategy of the institution.

Developing and implementing a data management policy and developing relevant services is essentially a team effort requiring the collaboration of a number of actors located within research institutions. The main units involved are the research office, the IT departments, the academic units, the libraries and the researchers. The research office (also known under other names) is the grant and contract administrator which acts as the liaison point between funding agencies and the research community within the institution. It provides support and guidance by essentially allowing researchers to comply with the funders' policies (which may also pertain to issues related to data management). Further support and guidance to researchers is provided by the departments/ units within which researchers are located and carry out their research and can involve assistance in proposal writing, compliance with funders' and/or university requirements. Horizontal coordination is provided through committees set up within institutions to facilitate policy development and the coordination of the flow of information among the various schools and departments.



[64]National Health and Medical Research Council, "Australian Code for the responsible conduct of research". http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf .

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

The policy defines the responsibilities of the researchers and the university, and highlights the importance of partnerships both within the university as well as with actors outside "to implement good practice and meet relevant legislative, research funder and regulatory requirements". The researcher is responsible for managing his/her research data according to university and funder requirements, while the university is responsible for providing researchers with access to services and facilities to meet their requirements, with training, support and advice. In addition, the university is responsible for providing the necessary resources to those operational units charged with the service provision.

A specific sub-committee of the University Research Committee was charged to guide the development of services and update the data management policy of the institution (Oxford policy, point 11). Additionally, a Working Group was appointed within the Committee to 'steer development and ensure senior management remain informed, as well as to ensure the involvement of all academic divisions'. The Group also helped with prioritization of requirements and advice on business models (J. Wilson, op cit. 2014. p. 5).

University of Oxford, Policy on the Management of Research Data and Records. 2012. http://researchdata.ox.ac.uk/files/2014/01/Policy_on_the_Management_of_Research_Data_and_Records.pdf



The role of the university/research library is central in the development of data management services as libraries have become very active in the field of open access to research publications, both in terms of advocacy and training, as well as in running the repository services of universities and research institutions, in collaboration with the IT departments or on their own, when the size of the institution and the technological capacity of the library permits it. With reference to research data, libraries are becoming involved in data curation, often acquiring expertise and offering data repository and other archiving and management services to researchers. Mostly, however, libraries are the main training, support and advocacy partner in data curation on behalf of the institution. Research libraries are close to researchers and able to scope their needs, develop services that suit them, on the basis of their particular disciplinary needs, provide training and support to researchers in many levels, such as data management planning, legal and copyright issues, technical standards for research data.[65]

When it comes to developing services to support research data management, 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, and the training for the researchers and advocacy services. It is common that when IT departments are involved they undertake the software and infrastructure development, while the library supports archiving, training and advocacy activities.

[65]Cf. LIBER, Ten Recommendations for Research Libraries to get started to research data management, 2012, http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf .

**Developing Data Management Services at the University of Oxford**

In developing data management services, the relevant departments involved are the Bodleian Libraries, IT Services and researchers from the Department of Zoology. At Oxford, services related to the planning and the bidding phases of the research data lifecycle are overseen by Research Services, services related to data management during the 'live' phase of a project are managed by the IT services, and services related to the post-project data curation, preservation, discovery and access are managed by the Library services (J. Wilson, op. cit.2014. p. 2-3).

The development of data management services requires institutions to consider which services to develop in-house and which to outsource, following an assessment of their needs and resources. With respect to infrastructures, while they are generally more developed in comparison to the associated policy frameworks, dedicated research data repositories are not widespread among research institutions.[66] In some cases, such as Oxford and Edinburgh, institutions may choose to develop a fully-fledged infrastructure and services to support researchers in managing research data. There are cases, however, where institutions instead of developing an institutional repository choose to collaborate in sharing resources for research data management. This is achieved through the development of an institutional registry that tracks where the research data of the institution lies. Collaboration in the development of the relevant infrastructures (primarily with data centres) contributes to capitalizing on existing investments and expertise in a distributed way, and towards scalability, even at national level.



**Multiple stakeholder collaboration for research data services in the Netherlands**

In the Netherlands, RDNL, an alliance of DANS, 3TU.Datacentrum and SURFSara, provides data services of national scale. The partners include data centers (DANS), high-computing centers of expertise (SURFSara) and the consolidated university service for research data(3TU) of the three major technical Dutch Universities, Delft, Eindhoven and Twente. This collaboration has the possibility for scaling up services by introducing other partners and providing nation-wide services. The RDNL will conclude institution-wide framework agreements for data management and storage services. They may involve division of responsibilities, as well as the establishment and maintenance of technical infrastructure.
Source: http://www.researchdata.nl/

[66] Science Metrix (Open data policies and strategies), op. cit.

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, since 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. The funding challenge is pressing even in the case of institutions like CERN where the provision of resources for the preservation of data is not evident.[67] Case studies on the costs of developing data management services indicate that this is a learning process, whereby institutions are able to pinpoint exactly the costs on the basis of their learning experience in implementing pilot services (e.g. University of Edinburgh).[68] At Oxford, for example, services that are considered core to the function of the University are supported by it (obligatory services), whereas each separate service within data management that may not necessarily be a core university service, may require the exploration of alternative sources of financing:[69] this can be achieved either by charging for these services or through establishing collaborations for creating economies of scale.

In terms of training, research performed by RECODE shows that 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 to a culture of open science more generally.[70] Researchers in some fields may require training because they lack the knowledge and skills on making their research data available and accessible, or on re-using data and incorporating data in their research process. Librarians and information experts require more training in providing research data services that are necessary in an increasingly data-intensive research environment. Thus, workshops, as well as more 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. The data-science initiative at New York University is one such initiative aimed at meeting the demand for skilled researchers and professionals in automated methods for analyzing data.[71]Similarly, for librarians, technical staff and information professionals, programmes for the development of skills are necessary so that they can, in turn, cover researcher needs. One such example is the specialization in data curation offered by the Graduate School of Library and Information Science of the University of Illinois.[72]



In the context of a research culture that is generally not conducive to fast changes, research institutions need to address the issue of cultural change, which will ultimately legitimize research data as important research outputs for professional advancement. This is a challenging and complex task since variability in research practices between disciplines has led to some being more open than others towards data management and data sharing.[73]Data-driven disciplines have, understandably, made more strides towards open data sharing and open

[67] Noorman et al, op. cit. 2013.
[68]LERU, LERU Roadmap for Research Data, Advice Paper, No14, December 2013.
[69]Wilson, op cit. 2014.p. 3.
[70]Noorman et al., op. cit., 2013, pp. 42-51.
[71]Data Science at NYU, http://datascience.nyu.edu/profession/ .
[72]http://www.lis.illinois.edu/academics/degrees/specializations/data_curation .
[73]Diane Harley, Sophia Kryzs Acord, Sarah Earl-Novell, Shannon Lawrence, C. Judson King Assessing the future landscape of scholarly communication. An exploration of faculty values and needs in seven principles. Centre for Studies in Higher Education. UC Berkeley: Centre for Studies in Higher Education. 2010.

science, while others, such as the humanities, have progressed less.[74]Awareness-raising and advocacy initiatives will contribute toward reducing the hesitancy of researchers to open and share their data, which is based on competition for prestige and funding and the fear of being scooped.[75]

An obstacle in the current research culture and professional environment that requires attention is the fact that currently there is little, if any, formal recognition for data outputs in academic promotion or other assessment processes. Indicatively, even in the UK, which is considered a pace-setter within the EU in terms of promoting open access, HEFCE excludes data from its policy on open access in the post-2014 REF.[76] Further progress is certainly to be expected, and will become apparent as a result of internal deliberations within research institutions and collaboration between the administration and researchers, as well as through disciplinary approaches, for example greater awareness-raising in scholarly societies and research communities that will enhance the bottom-up process. In such a context the Wellcome Trust Report[77] recommends the explicit recognition and assessment as research outputs of high quality datasets in the post-2014 REF. Eventually, institutions need to consider the ultimate incentive for researchers, which is the formal acknowledgement of research data as legitimate research outputs that count towards their promotion.

Finally, RECODE research clearly demonstrated that legal and ethical challenges that researchers and institutions are faced with in dealing with open research data need urgent and systematic attention.[78] Researchers face numerous challenges in this front while performing research, producing research and data, as well as when accessing them for reuse, and they lack systematic assistance and guidance for the most part. Further, open access requirements by funders may often come into contrast with existing legal and other contractual obligations and frameworks. Legal and ethical issues raised by improving access to increasing volumes of research data, thus, requires careful treatment horizontally at the institutional level, as well as under disciplinary lenses through departmental and other instruments.



### *Recommendations*

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

   Research Institutions should address research data and develop explicit relevant strategies and policies. These should define the roles and responsibilities of the different stakeholders involved in implementing them (i.e., individual researchers, university libraries, research committees, IT departments etc.); they should state the institutions' responsibility in seeking appropriate funds from funding bodies to develop the appropriate capacity (infrastructure and services and monitoring tools). In doing so, consultation and collaboration with the research community is of critical importance in understanding the latter's needs and in developing the necessary infrastructure and services to meet these

[74]Overall as supported by the RECODE case-studies while researchers are positive about open access they remain quite skeptical in relation to the practicalities involved (Sveinsdottir et al. op. cit., 2013, p. 74); the variability in data-sharing attitudes between different fields is also clear in Van den Eynden, V. and Bishop L. (2014). Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report http://www.knowledge-exchange.info/Default.aspx?ID=733

[75]Sveinsdottir et al., op. cit. 2013

[76]Higher Education Council for England. "Open access in the post-2014 Research Excellence Framework". http://www.hefce.ac.uk/pubs/year/2014/cl072014/#d.en.86764

[77]The Wellcome Trust, Establishing Incentives, op.cit.

[78] Finn et al 2014, op. cit.

needs. In developing related policies, the establishment of committees within institutions that are expected to work in close collaboration with funders and the research communities, to explore and assess different approaches in data sharing will alleviate significant pressure from researchers and accommodate disciplinary practices. Attention should be directed towards providing forums, as well as formal processes and procedures (for example through specific ethics and other relevant committees), where legal and ethical issues arising from opening up research data can be dealt with effectively (e.g. licensing frameworks).

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***
   Implementing a strategy for open access to and long-term preservation of research data requires leveraging considerable resources and developing services, infrastructures, training and support systems, among others. It further requires significant financial investments sustained over the long term, the latter usually being taken up largely by research institutions. To achieve a sustainable ecosystem and infrastructure for open access to research data and their long-term preservation, collaboration between and within institutions is critical. Developing relevant services requires the collaboration of different institutional departments within an institution, for example high administration, IT services, library and specific departments, among others. 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***
   To unleash the potential of research data for further exploitation by the research community, businesses and the public, research institutions are encouraged to formalize the changes already underway in the scholarly communication system towards greater openness by including data sharing of high quality research data as a formal criterion for career progression. Such formal recognition should be accompanied by the development and use of metrics that allow the collection and tracking of data use and impact. Advanced career professionals (researchers and professors) are encouraged to lead by example as well as to help address this issue formally with scholarly societies and associations and, in particular, within their research institutions and departments. Formal acknowledgement of research data as a legitimate output will incentivize researchers and bring gradual change in practices.

4. ***Develop educational and training programmes for researchers and staff to improve data management skills and to enhance data-intensive research***
   Researchers are expected to acquire knowledge and skills on working with various formats and software tools, mark-up languages, meta-data etc.,[79] in addition to familiarizing themselves with legal and ethical aspects related to open access. In designing such programmes research institutions should pay attention to disciplinary specificity and practices. Training on technical requirement is expected to be provided by IT departments, while training on data management and the discovery of data by university

[79]Noorman et al., op. cit., 2014 p. 42.

libraries.[80]Training programmes for librarians and data scientists/ information specialists will allow them to be kept informed of the latest developments, be able to meet researcher needs and contribute further to the development of data science. In doing so, research institutions- with the university departments assuming a leading role- can explore the possibility of developing joint courses with data managers who have the expertise in data management, especially data centers, and across different specialties. Research institutions are finally encouraged to develop curricula that recognize data science and information management as distinct career paths. In developing such programmes, institutions should assess their capacities in meeting their preservation and sharing goals, while recognizing disciplinary diversity and thus avoiding "one-size-fits-all" solutions.

5. ***Raise awareness about the benefits of open access to research data and provide rewards***
Cultural change is a difficult challenge not only because it takes time but most importantly because it has to address the reservations of the researcher community to share their data due to the fear of being scooped, as a result of the competition within science for career and reward. Focusing on awareness—raising and advocacy activities, as well as creating incentives for 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. Such activities can be directed at engaging researchers with varying levels of engagement in sharing research data, and can be either general in nature, or targeted to specific disciplinary communities. Institutions, along with funders, might also wish to consider offering small grants to incentivize data sharing. This would not only give researchers a further incentive but might also engage their curiosity and generate examples of good and innovative practice.



6. ***Support the research community through 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. These should be addressed at the institutional, disciplinary and/or technical level.

[80]Ibid.

## 2.3 DATA MANAGERS

### *Context*

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. It refers to data centres, which are mainly government financed operations for making datasets available, like the Swedish National Data Service or (national, university or specialized) libraries, archives and memory institutions that maintain collections of content.[81] Some of these institutions, traditionally charged as content gatekeepers, have developed strength in relevant technological infrastructures for the storage, curation and long-term preservation of digital data, while others are still lagging behind. This analysis focuses mostly on data centers and libraries as key stakeholders in fostering change.

Data centres come in different forms and sizes and even though they often emerge from a disciplinary community, their success is built on their openness to the global community and their commitment to working with other stakeholders in that community. Data centres are more widespread in disciplines where a culture of data sharing is more established, such as molecular biology, earth observation and some social science disciplines, and less so in areas where an open data culture is still underdeveloped.[82] The most basic function of data centres relates to the storing of research datasets for a defined community and making them accessible for other researchers to discover and use. In essence, this entails two roles: firstly, ensuring that data is discoverable by providing a single location where researchers can deposit their work, as well as 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.[83]



**The National Oceanographic Data Centre:**
**A national disciplinary data center and its services**

The NODC is one of the national environmental data centres operated by the National Oceanic and Atmospheric Administration (NOAA) of the US Department of Commerce, which manages the world's largest collection of publicly available oceanographic data. According to its charter, the centre 'serves to acquire, process, preserve, and disseminate oceanographic data'.
Its primary mission is 'to ensure that global oceanographic data sets collected at great cost are maintained in a permanent archive that is easily accessible to the world scientific community and to other researchers'. In fulfilling its mission it collaborates closely with other national and international organizations such as the Global Earth Observation System of Systems (GEOSS), and the International Oceanographic Data and Information Exchange (IODE) dedicated to open access data.

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

[81]The PARSE.Insight project defines data managers as profit and non-profit data archives, traditional memory institutions (libraries, archives and museums) as well as research and development in preservation technology. Van De Hoeven, J. Insight into digital preservation of research output in Europe, PASRSE.Insight, D3.6, 2010, p. 11-12. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf

[82]ODE *Summary of the studies, thematic publications and recommendations.*D6.1. 28 October 2012. p. 15. http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/11/ODE-WP6-DEL-0001-1_0.pdf

[83]Ellen Collins, Use and Impact of UK Research Data Centres, The International Journal of Digital Curation, Issue 1, Volume 6, 2011, pp. 20-31. doi:10.2218/ijdc.v6i1.169

**Data Archiving and Networking Services (DANS)**

DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), and is also supported by the Netherlands Organization for Scientific Research (NWO). Since its establishment in 2005, DANS has been providing storage of and continuous access to digital research data in the social sciences and humanities.

DANS offers a vast number of services to ensure that access to data keeps improving. These services include, but are not limited to the following: certification of repositories through the provision of the Data Seal of Approval; data management plans; funding of small data projects focusing on describing and making datasets accessible or carrying out feasibility studies; the EASY data archiving and online access to data.

Source: http://www.dans.knaw.nl/en

Libraries are also repositioning themselves in the context of the emerging digital science environment. Traditionally providing access to resources and publications to researchers through subscription, libraries are finding themselves actively seeking and acquiring a new role: that of the custodian of the research publications and research data of their institution. Being active in open access, libraries usually operate institutional repositories and are strong open access advocates. 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, the RECODE confirmed[84] the findings of ODE project[85] that libraries fall short both in terms of their current practices (as most libraries still lack a digital preservation strategy), 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 -specialized disciplinary data centre or research library- data managers should address open access to research data as a current and 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. For example, the Digital Curation Centre has developed a useful and detailed guide to assist institutions to develop data management services.[86]

[84]Noorman et al. op. cit., 2013.
[85]ODE, op. cit., 2012.p. 29.
[86]Jones, S., Pryor, G. & Whyte, A. 'How to Develop Research Data Management Services - a guide for HEIs'.2013. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/how-guides - See more at: http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services#sthash.jLdE7VJW.dpuf

Common collaborations are those with publishers, facilitating the smooth linking of publications to research data; with funders, by offering them trusted locations for storing and providing access to research data produced from the research they fund; with researchers and scholarly associations in developing disciplinary metadata standards. Collaboration for shared data services are extremely important in achieving economies of scale in infrastructures and services at the national and international scale as the example of RDNL demonstrates.[87]

**Funder-Data Center partnerships: UKDS and UKDA**

A close partnership between a Data Center and a funding institution can be observed in the case of the UK Data Service. The UKDS was established in 2012 with the support of ESRC to integrate and expand several existing data services into a comprehensive national service structured to support researchers in academia, business, third sector and all levels of government. It is supported by experts in various Universities across the UK.

The ESRC has also provided a long-term commitment of funds for the UK Data Archive (UKDA), which has been central to its success in providing access to research data across for the entire country.

Sources: http://ukdataservice.ac.uk/ and http://ukdataservice.ac.uk/



The costs of data management and curation services are an issue of great importance for data managers. 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.[88]Currently, many data centres and repositories receive their funding for open research data primarily from public and private funding bodies, either directly (for example the ESRC and NERC support specific data centres) or through grant funding, while subject data repositories may be developed as part of research projects. Reliance exclusively on project funding is nonetheless problematic, as it does not always guarantee long-term funds and, thus, operations. For example, the withdrawal of its single source of income led the Arts and Humanities Data Service in the UK to close down its operations,[89]raising questions as to the future of hosted research data. Thus, developing sustainable funding models on the principles of diversifying sources of income and establishing collaborations should be addressed with particular care. The 4c project, an FP7 funded project, is addressing the cost of digital curation to help European organizations in investing more effectively in digital curation and preservation.[90]

[87] Research Data Netherlands http://www.researchdata.nl/en/

[88]Beagrie, Neil, Brian Lavoie and Matthew Woollard, *Keeping Data Safe (Phase 2),* Jisc,30 April 2010. Noorman et al., op. cit., 2013

[89]Dallmeier-Tiessen, Sunje, Robert Darby, Kathrin Gitmans, Simon Lambert, Jari Suhonen, Michael Wilson, Compilation of Results on Drivers and Barriers and New Opportunities, 2012. http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2012/08/ODE-CompilationResultsDriversBarriersNewOpportunities1.pdf .

[90]The 4c's project draft Roadmap "Investing in Curation: a shared path to sustainability" is available file://filesrv/userdocs2/mangelaki/My%20Documents/Downloads/D5.1%20Draft%20Roadmap%20-%20V1.01%20-%2008Aug2014.pdf .

A further important contribution of data managers is towards the proliferation of high quality research data, in other words, in 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.

**Cost-models for data services**

**DANS** charges basic data storage costs, including backup. Storage is supplied by an external party, which guarantees the high level of security and data availability. In return for one-off payment of these costs for five years in advance, DANS ensures conservation of the data 'forever'. This enables long-term preservation of data funded with temporary project subsidies. Formatting data and consultancy is charged.

**3TU.Datacentrum** provides services at different prices to partners, contracted customers, and customers without a contract. All pay a one-off fixed rate per TB with options for retention periods with Datacentrum partners paying less that the other two. No fee is charged to researchers outside of these three categories for uploads up to 4GB. The Datacentrum has a service level agreement with TU Deltf's Shared Service Center for reliable storage and backup.

Source:
http://www.researchdata.nl/fileadmin/content/RDNL_algemeen/Documenten/RDNL_FOBOmodel-UK-web.pdf



Several repositories and data centres have developed quality assurance measures and offer a range of services to evaluate the technical quality of data sets.[91] 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 altmetrics.[92]

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. The former is assessed on the basis of criteria such as formats, metadata, DOI etc. The latter, i.e. scientific quality is more a task for the research community through the definition of appropriate scientific criteria. Nonetheless, data managers can ultimately play a role in this by accepting publications which abide to the criteria set by the scientific community.

Aside from the quality of the research data, the quality of services offered by data centres and repositories is becoming a cutting edge issue. As shown in a related survey, while "the amount of digital data explodes and a growing number of institutions are establishing digital archives, there is still deficit in standards and commonly accepted measures used for the development

[91] Noorman et al, 2-13.
[92] Ibid.

and the quality control during the creation of such archives."[93] 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.

**The Data Seal of Approval**

The DSA is a certification awarded to trusted and sustainable repositories on the basis of fulfilling relevant guidelines provided by the DSA and following review. This provides assurances to patrons regarding the reliability of the services. DSA was developed by DANS and has operated under an international board since 2009.

Source: http://www.datasealofapproval.org

The transition to an open access environment requires researchers to know how to work with various formats and software tools for making their data publishable and reusable as well as how to search for the data they need. It also requires library personnel and other staff to keep up with technological developments that will allow them to support researchers in an effective way. This essentially points to the need for investment in skills and training development, an area where intensive work is required.[94] Data managers have a significant role to play by providing training to researchers for meeting technical quality standards with their data sets, as well as in developing disciplinary standards. Additionally, data centers with expertise in data curation have an important role in helping improve the skills of research library staff in data managements, data quality and developing data services.



**Providing Training and Expertise: The Digital Curation Centre**

DCC is a world-leading centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management across the UK's higher education research community. DSS provides a wide range of resources and guides for developing data management services and has been instrumental in providing training for researchers and data managers in particular in the UK, as well as in the rest of Europe.

Source: http://www.dcc.ac.uk/

[93] Dobratz, Sussane, Peter Rodig and Uwe M. Borghoff, Bjorn Ratzke, Astrid Shoger, *The Use of Quality Management Standards in Trustworthy Digital Archives*, The International Journal of Digital Curation, Issue 1, Volume 5, pp. 46-63. http://www.lis.illinois.edu/academics/degrees/specializations/data_curation .
[94] Noorman et al., op. cit., 2014.

***Recommendations***

1. ***Assess their position within the open access ecosystem in view of developing collaborative infrastructures and services***
   Research libraries are encouraged to evaluate their infrastructure and services for research data management and assess how these can serve their designated community in the best possible way. This will allow them to evaluate whether to proceed in developing them further, outsource some of them or provide shared services (e.g. infrastructures). This process should take place in close collaboration with the administration of the institution, since data management services complement wider institutional policies on data management and with the research community they serve. Collaboration with other stakeholders, such as other universities and/or data centers is desirable, in view of establishing scalable and sustainable infrastructure and services, and swiftly acquiring know-how that exists in external institutions. Data centers are also encouraged to collaborate with other stakeholders (e.g. research institutions, publishers and funders) in developing new, improving existing services and infrastructures or sharing some elements of their services that address the needs of the research community.

2. ***Develop sustainable business models to ensure long-term service provision***
   While a common strategy for the development of services has been through funding bodies or grant funding the need to make considerations for long-term and reliable service provision points to the significance of developing sustainable business models. 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. For universities and university libraries or data centers this requires the involvement of the administration, as some of these services may incur a regular expense that the research institution is called to support as a core service to its faculty and students. Multi-institutional collaborations are crucial in establishing sustainability and scalability of services.



3. ***Establish mechanisms for data quality that ensure re-use and long-term preservation through collaborative work***
   Institutional repositories, data centres and publishers have a responsibility to ensure the quality of metadata, file readability and adherence to standards, along with the development of policies for long-term preservation and re-use of data. To fulfil it they have a range of quality assurance and control strategies to use, both manual and automatic. While data managers are expected to take the leading role in close collaboration with research communities in establishing citation standards, their collaboration with publishers and journal editors (in essence through editorial policies) is central in ensuring their further enforcement. Data managers are also encouraged to provide support to research communities in developing strategies evaluating the scientific quality of data, for instance through online platforms, where research communities can discuss and rate data sets or link data to scientific publications. The quality of research data quality is increasingly becoming important in decisions regarding data retention, as data volumes grow exponentially.

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 also for the data centres hosting them. Data centres are therefore 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***

   University libraries are the primary loci where researchers seek advice and support for data management and DMPs. 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. Furthermore, training for researchers is necessary for intellectual property rights, licensing, re-use of research data, advising on choosing deposit loci external to the university, among others. Collaboration between libraries and data centers may help the former acquire skills and provide specialized training to their researchers. Cooperation between data managers and research communities is of key importance in allowing the former to understand the data management needs of the latter and, thus, in developing appropriate disciplinary-specific training and standards. Finally, data managers, especially experts within data centres, should work on training courses and advising research institutions on curricula both for researchers, as well as for library staff, in view of further developing the skills necessary to work with increasingly data-intensive research.

## 2.4 PUBLISHERS

### *The Context*

As a result of the journal-based dissemination structure of research, publishers are key stakeholders in the open access ecosystem.[95] They are, thus, in a unique position, in cooperation with funders, data centres/repositories and research institutions 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 from very small non-profit scholarly led, university-based operations, and small entrepreneurial ventures, to giant multinational enterprises that overwhelmingly control the market.[96]

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 (i.e. publications) thus enhancing the trustworthiness of the published research and the ability to verify it, which lies at the heart of the ethical conduct of research.[97] Additionally, funder policies requiring open access to research data essentially adds pressure to publishers to engage with open access to research data in order to help authors and institutions comply with funder requirements. Finally, the recent attention to research data is gradually leading publishers to exploit the possibilities of research data in new data-based products and services, such as the data journals, peer-review to research data, and services to enhance data quality.



Scientific, Technical and Medical (STEM) publishers working in data-intensive fields were the first to acknowledge the significance of open access to research data. STEM publishers endorsed the Brussels Declaration of International STEM principle that "Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars"[98]. Publishers, mainly from the STEM fields, have recently turned to developing mandatory policies requiring authors to deposit research data underpinning their publications in certified repositories and make them openly available. However, a 2013 survey conducted by the Journal Research Data Policy Bank (JoRD) shows that only 31 out of a 371 journals possessed policies for open access to research data supporting publications.[99] A recent study indicates that for journals with a combination of mandatory data policy and an accessibility statement the odds of finding the data online were almost a thousand times higher compared to journals with no policy.[100]

[95]Sveinsdottir et al., op. cit., 2013.
[96] On the ever-growing market, especially in STEM publishing, cf. press release on Simba report for 2012, where this market is estimated to have grown 3.4% to $21.1 billion in 2011. http://www.simbainformation.com/about/release.asp?id=2503
[97]The Royal Society, op. cit., 2012.
[98]STM, "Brussels Declaration", 1 nov., 2007, http://www.stm-assoc.org/brussels-declaration/#.
[99]Jord, "going back to basics – reusing data", blogpost, July 5, 2013. http://jordproject.wordpress.com/
[100]Vines, Timothy H., Rose L. Andrew, Dan G. Bock, Michelle T. Franklin, Kemberly J. Gilbert, Nolan C. Kane, Jean-Sebastien Moore, Brook T. Moyers, Sebastien Renaut, Diana J. Rennison, Thor Veen and Sam Yeaman, "Mandated data archiving greatly improves access to research data, *The FASEB Journal*, article fj. 12-218164, published online January 3, 2013.

**The PLOS mandatory policy for open access to research data**

PLOS instituted a mandatory open access policy for all research data supporting PLOS publications with rare exceptions. Researcher compliance with PLOS's policy is stated in the Data Availability Statement submitted by the author simultaneously with the submission of the manuscript, and published as part of the final article in cases of successful submissions. Authors are required to deposit their data in recommended certified data centers and repositories. Small data sets can be uploaded data with manuscripts. PLOS' Editorial and Submission Policies provide support, guidance and assistance provided to researchers to allow them to comply with the journals' open data policy. In cases where restrictions to access are noticed following publication PLOS reserves the right to post a correction, contact the authors' institution or funders or even retract the publication.

Source: PLOS One.http://www.plosone.org/static/policies.action#sharing

The increasing interest of the publishing community to open data should also be assessed in light of the emergence of a new publishing product, the data journal. Its emergence should be linked to the efforts of publishing data separately "to ensure that these essential parts of the scientific record were made available in an intelligible form to the scientific community".[101] 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. As pointed out by ANDS in its Data Journal Guide 'Fundamentally, data journals seek to promote scientific accreditation and re-use, improve transparency of scientific method and results, support good data management practices and provide an accessible, permanent and resolvable route to the dataset.' Again as highlighted by ANDS, 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. [102] Finally, as data papers are becoming distinct publishing products, a number of data journals are also supporting alternative metrics (altmetrics),[103] thereby enhancing further data publication.



[101]International Council for Science (ICSU). Open access to scientific data and literature and the assessment of research by metrics. p. 8 .
[102]Australian National Data Service. "Data Journal Guide" http://ands.org.au/guides/data-journals.html .
[103]Altmetrics is the study and use of scholarly impact measures based on activity in online tools and environments.

**The Geoscience Data Journal**

The Geoscience Data Journal is an online journal published by Wiley. It publishes short data papers which are linked to their relevant data sets. The journal collaborates with a number of data centers and repositories, where the relevant data sets can be deposited by authors. The latter provide trusted loci to store data and link to via a DOI. Data creators attain full credit for their efforts, while also improving the scientific record by providing access to research data that is fully described, edited and discovered.

Source: The Geoscience Data Journal

**The Journal of Open Archaeology**

The Journal of Open Archaeology (JOAD) applies a peer review process to all submitted data papers against two criteria: the paper content and the deposited data. According to the journal, the former is about providing information regarding the creation and re-use of the dataset as well as a description of the dataset, while the latter is among others about the submission of data to a repository with a sustainability mode, its licensing

Source: http://openarchaeologydata.metajnl.com/
Source: http://openarchaeologydata.metajnl.com/about/editorialPolicies#peerReviewProcess



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. Against this background, several publishers have expanded peer review to cover data as well. According to Mayernik et al. (2014) "data peer-review" and the processes used can vary by the kind of publications or resources being reviewed.[104] Thus, they make a distinction between: data analyzed in traditional scientific articles, data articles in traditional scientific journals, open access repositories and datasets published via articles in data journals. Notwithstanding these differences, Mayernik et al. (2014) have identified some common elements in relation to data peer review and the quality assurance process of data, namely: the need for data accessibility (via data centers or repositories); the provision of adequate information for the dataset to be reviewed; clear guidelines for data peer-reviewers on how to perform the data review and what characteristics to be examined.[105] 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.

[104]Mayernik, Mathew S., Sara Callaghan, Roland Leigh, Jonathan Tedds, Steven Worley, "Peer Review of Datasets: When, Why and How", *Bulletin of the American Meteorological Society*, Early Online Release, May 7 2014. http://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-13-00083.1 .
[105]Ibid.

**The FORCE11 Data Citation Principles**

To enhance the proliferation of high quality research data, contribute towards according research data due significance in the research process and encourage good practice, FORCE11 issued overarching principles for data citation. The principles work on the premise that data citations need to be both human and machine-readable. They are not comprehensive, but are rather meant to encourage communities to develop practices and tools that embody them.
The principles are: importance of data citations; data citations to facilitate credit and attribution; data citations to provide evidence for claims; data citation to include machine actionable and globally unique persistent identifier; data citations to facilitate access to the data themselves. In addition, data citations should be: persistent; facilitate identification of, access to, and verification of the specific data that support a claim; be interoperable and flexible.

Source: https://www.force11.org/datacitation



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. As stated in the Report from the Expert Group on Text and Data Mining 'TDM is an important technique for analyzing and extracting new insights and knowledge from the exponentially increasing store of digital data ('Big Data')'.[106] The same report concludes that TDM is expected to gain further importance as researchers will increasingly acquire the skills and technology for investigating datasets of increasing size, complexity and diversity. As open access publishing is growing it provides further support to those arguing that the use of TDM should not be faced with restrictions.

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.[107]This policy seems in sharp contrast to the recent UK legislation removing licensing restrictions to TDM.

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

[106]European Commission, Standardization in the area of innovation and technological development, notably in the field of Text and Data Mining. Report from the Expert Group. 2014 http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf .

[107]A typical example showing the difference of approach as to what is considered unobstructed TDM can be found in Elsevier's recent TDM policy and the criticism this received by a total of eighteen European research and library organizations. In particular, the above organizations have called on Elsevier to withdraw its current policy arguing that it places unnecessary restrictions on researchers, ultimately limiting their ability to mine content to which they have legal access. Cf. LIBER. "European research organization call on Elsevier to withdraw TDM policy". http://libereurope.eu/news/european-research-organisations-call-on-elsevier-to-withdraw-tdm-policy/ .

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 editorial principles that promote citation of research data through the development of disciplinary-specific standards alongside internationally accepted principles (e.g., FORCE11), as well as data review processes.[108] 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***
   Publishers should support their journal editors in developing mandatory policies for research data as a means of increasing transparency in the research process and enhancing data research. Editorial policies should address issues like documentation, metadata and format of published data, licensing, 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, such as repositories with a Data Seal of Approval. Such repositories or data centers can be subject-specific, institutional or general repository services. Their collaboration should aspire to offer operationally and technically seamless high quality products and services to researchers and research institutions that lead to the dissemination and preservation of high-quality scholarship and research data in open formats. 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. In that respect, publishers' role should be more about prompting the community to implement repository assessment guidelines thus helping researchers in choosing the appropriate repository and 'once [these] standards are established by the community publishers can then enforce them through journal policies'.[109]

3. ***Support data as a first-class scholarly output through the establishment of peer-review processes***
   Taking into consideration that one of the main roles of publishers is to ensure the publication of high quality research, 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

[108]Dreyer, Erwin "Open Data: A role for publishers… or for journal editors?, Plos Biology, http://ist.blogs.inra.fr/afs/2014/11/01/open-data-a-role-for-publishers-or-for-journal-editors/
[109]Jennifer Lin and Strasser Carly, Recommendations for the Role of Publishers in Access Data, PLOS Biology, volume 12, issue 10, October 2014

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. In adopting peer-review processes publishers should also explore ways of incentivizing reviewers; indicative measures include payment of honoraria, invitation to write special articles and join editorial boards or even employment of some experts as 'in-house' reviewers.[110]

***4. Develop policies requiring citations for research data***

The adoption of the traditional publication process for research data should be accompanied by the establishment of standard citation practices in view of enabling re-use in the long term. In that respect, publishers should require that data accompanying their publications are citeable, and provide clear guidelines for data citation. In developing citation policies, publishers may consult principles, such as the Force11 principles, become involved in the DataCite initiative and involve the research community and editors in the process. Data citation should include DOIs, as well as licensing information (e.g. Creative Commons licenses), preferably machine actionable, that lets users know what they are able to do with research data.

***5. Establish licensing policies that encourage the use of TDM***

Editorial policies should state in a clear way 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.



[110]Costello, Mark J., William K. Michener, Mark Gahegan, Zhi-Qiang Zhnag and Philip Bourne, *Biodiversity data should be published, cited and peer-reviewed,* Trends in Ecology and Evolution, Vol. 28, No 4, August 2013, pp. 454-461.

# ANNEX: RECODE GUIDELINES FOR POLICY IMPLEMENTATION

## 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, minimally, with research institution administration, researchers and particularly with 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.** Policies should be developed with open access as the default. These policies should be explicit and publicly available and address a number of recommended issues (cf. section below on policy content).
- **Data Management in grant applications.** Where data is generated, data management plans (DPMs) should form an essential element of the grant proposal, be resourced adequately, reviewed and their delivery monitored.
- **Provision for relevant open access clause in grant agreements.** Grant agreements should include clauses on open access to research data, accompanied by the description of sanction mechanisms in cases of non-compliance, as well as clarification of eligible costs.
- **Guidance to researchers.** Funders should develop appropriate tools such as templates for data management and resources on data management and DMPs, and specify the criteria for eligible repositories/data centers for data deposit.
- **Rewards to researchers.** Beyond establishing a mandatory policy, further measures 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 for 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 should 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. 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 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 and participation in international fora.
- **Participation in dialogue and collaboration** among stakeholders both within and outside the institution (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 assumes the responsibility to provide 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 for other data by the end of the project.
- ***Licensing.*** The policy should require research data to be accompanied by licensing describing the terms of use, such as Creative Commons. 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 publishers' collective instruments and research 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 the 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 publishers' collective instruments and research communities in addressing/ developing polices 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 polices?
- Have you developed indicators for measuring data impact from your publications?
- Do you encourage data sharing through specific actions, such as prizes?

## RESOURCES

### *For Funders*

**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/Horizon 2020 Rules for Open Access**

- **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 to support new types of data 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/



### *For Research Institutions*

**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

**Multi-institutional collaboration for research data management**

- **RDNL**
  http://www.researchdata.nl/

**Data Management Plans**

- **Guidance and resources for Data Management Plans**
  http://www.dcc.ac.uk/resources/data-management-plans

### *For Data Managers*

**Data Management Services**

- **Digital Curation Centre**
  http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-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

**Cost models for data services**

- **3TU.Datacentrum**
  http://datacentrum.3tu.nl/en/home/

**Partnerships (Funder-Data Centre Partnership)**

- **UK Data Archive**
  http://www.data-archive.ac.uk/

**Training and Expertise**

- **Digital Curation Centre**
  http://www.dcc.ac.uk/training



**Accreditation Service**

- **Data Seal of Approval**
  http://www.datasealofapproval.org

### *For Publishers*

**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/

**Collaboration (Publishers and Repositories)**

- **Geoscience Data Journal**
  http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%292049-6060

**Data Citation Principles**

- **The FORCE11 Data Citation Principles**
  https://www.force11.org/datacitation