

**European Landscape
Study of Research Data
Management**



SIM4RDM
SUPPORT INFRASTRUCTURE MODELS
FOR RESEARCH DATA MANAGEMENT

About this Publication

*European Landscape Study of Research Data Management
For SIM4RDM- Support Infrastructure Models for Research Data Management*

SURF
PO Box 2290
NL-3500 GG Utrecht
T + 31 30 234 66 00
F + 31 30 233 29 60

info@surf.nl
www.surf.nl/en

Authors

- Wilma Mossink - *SURF*
- Magchiel Bijsterbosch - *SURF*
- Joeri Nortier - *SURF*

SURF is the collaborative organisation for higher education institutions and research institutes aimed at breakthrough innovations in ICT (www.surf.nl/en)

This publication is online available through www.surffoundation.nl/en/publications

© Stichting SURF
May 2013

This publication is published under the Creative Commons Attribution 3.0 Netherlands Licence.



Contents

- Executive Summary..... 5**
- Introduction 7**
- 1 Methodology..... 8**
- 2 Desk research..... 10**
 - 2.1 Introduction 10
 - 2.2 Riding the Wave 10
 - 2.3 OECD Principles and Guidelines 11
 - 2.4 Data Management Policies 11
 - 2.5 Needs of researchers..... 14
 - 2.6 Data Management Plans 17
- 3 Analysis of the policies 19**
 - 3.1 Research Funders 19
 - 3.1.1 Policy for funding and data management 19
 - 3.1.2 Requirement of Data Management Plan 20
 - 3.1.3 Data preservation 21
 - 3.1.4 Reservation of funds for support and storage of research data..... 22
 - 3.1.5 Length of time that data is required to be made accessible 22
 - 3.1.6 Length of time that data is required to be preserved 22
 - 3.1.7 Evaluation of the Data Management Plan 22
 - 3.1.8 Experience..... 23
 - 3.1.9 Requirement that best advances research data management 23
 - 3.1.10 Conclusion 24
 - 3.2 National bodies..... 25
 - 3.2.1 National body for data management 25
 - 3.2.2 Funding of the national body 25
 - 3.2.3 Maturity of the national body 25
 - 3.2.4 Inspiration from across the border 26
 - 3.2.5 Tasks of the national body 26
 - 3.2.6 Conclusions..... 27
 - 3.3 Institutional policies 28
 - 3.3.1 Policies and data management 28
 - 3.3.2 Recognition..... 31
 - 3.3.3 Institutional research data support..... 33
 - 3.3.4 Institutional tools and analysis 36
- 4 Publishers..... 41**
 - 4.1 Research methods and response..... 41
 - 4.2 Summary of findings 41
 - 4.3 Conclusions..... 42
- 5 Validation of interventions 43**
- 6 Workshop findings..... 47**
- 7 Conclusions 49**
 - 7.1 Funding organisations 49
 - 7.2 National bodies..... 49
 - 7.3 Research institutions 49
 - 7.4 Publishers 50
- 8 Recommendations 53**
 - 8.1 Funding organisations 53
 - 8.2 National bodies..... 53
 - 8.3 Research institutions 53
 - 8.4 Publishers 54
 - 8.5 Recommendations from the workshop 54

Executive Summary

SIM4RDM is a 24-month project funded under the Seventh Framework Programme. The project aims to enable researchers to effectively utilise emerging data infrastructures by ensuring that they have the knowledge, skills and support infrastructures necessary to adopt good research data management methodologies. The aim of this study is to produce an overview of possible interventions that have proved effective in supporting researchers in their data management.

Based on the target areas identified by the Jisc Managing Research Data Programme model and desk research, an online survey has been conducted with a response of 18 national bodies, 20 research funders, 107 research institutions and 7 publishers. The findings of the survey and desk research were presented at a two-day workshop attended by 23 participants spread across the different stakeholder groups. Finally, based on the results of the previous activities a total of 12 researchers were interviewed to validate the effectiveness of interventions that were identified in the previous activities.

This report presents the results of an online survey to establish which interventions are already being used by funding agencies, research institutions, national bodies and publishers across the European Union member states and a number of countries outside Europe in order to improve the capacity and skills of researchers in making effective use of research data infrastructures. It also makes recommendations that organisations can adopt to help their researchers.

Various national bodies have been established to coordinate research data management activities such as access to tools and materials, in order to establish long-term preservation and sharing practices, assist in the policies of research funders and create infrastructures. Most of these national bodies are funded by the government, more often than not in a project structure. National bodies could take the lead in drafting a national code of conduct which encourages the creation and use of data management plans. They should also suggest and supply appropriate tooling and adapt these to the national context. Furthermore they should take an active role in data citation practices.

Interviews with researchers indicate that the main drivers for writing a data management plan are requirements by the funder or the publisher. Nearly half of the research funders who took part in the survey have a policy covering research data management, whilst a quarter of the funders require data management plans as part of the grant application. Data management plans should address data acquisition, use, re-use, storage and protection and the rights of ownership. Just over one third of the responding funding organisations designate a specific organisation for preservation, although no term has been identified.

Funding bodies should encourage researchers by offering clear instructions to create a data management plan at the level of the project proposal. Call funds should be allocated to research data management. Funders can designate centres to store research data.

For publishers, policies have yet to become established instruments. Policies that do exist, require readers and reviewers to provide links to the data underlying the article. In some cases publishers also require them to make entire datasets available when submitting the article, but not to keep them up to date. Publishers generally do not employ standards for data. Digital object identifiers are indeed used for citing, but data are cited in different ways.

A dialogue should be established with publishers and publishers' associations about the definition of data policies. Possible elements are persistent identifiers for citation of data and requirements of reliability for repositories in which data are to be deposited.

The number of research institutions with a data management policy is growing. Of those without a policy, 42% intend to define a policy within the next year. The main driver for the creation is the requirement by funders of a national code of conduct. For institutions without a policy, academic demand or general commitment to open access can drive the establishment of a data management

plan. 15% of the research institutions that have responded to the survey actually prescribe a data management plan. 50% of the responding organisations offer an infrastructure for storage and management of and access to research data made up by a variety of file storage and library systems. Research institutions should develop policies which contain elements that support scholars and scientists in their management.

Interviews with researchers show that policies should primarily cover roles and responsibilities for managing data, mechanisms for storage, backup, registration, deposit and retention of research data, access to re-use of data, open accessibility and availability of data and long term preservation and curation. They should also address data infrastructures and create workflows for data publishing and archiving. In addition, such policies can bridge the gap between data and publication by crediting researchers for publishing research data. Researchers indicated they would significantly benefit from face-to-face support and training, but they indicated that flyers, meetings, seminars and presentations are not very effective.

The setting for research data management is broader than initially anticipated. Other stakeholders need to be brought in as well, e.g. editorial boards of scientific journals, data centres and infrastructure providers. Research societies may intervene with the development of common practices. Infrastructure providers could intervene with common data formats for preservation and storage, tools and utilities. Policies from funders, institutes and editorial boards may influence researchers to use the principle of 'share and share alike'.

Introduction

SIM4RDM stands for Support Infrastructure Models for Research Data Management. SIM4RDM is a 24-month European project funded through the European Seventh Framework Programme under the theme 'Cooperation for ERA-NET supporting research structures in all Science and Technology fields'. The project team is made up of six partners from different parts of Europe: Jisc (United Kingdom), HEA (Ireland), NIIF (Hungary), Nordforsk (Norway), CSC (Finland) and SURF (Netherlands).

The aim of SIM4RDM is to enable researchers to effectively utilise emerging data infrastructures by ensuring that they have the knowledge, skills and support infrastructures necessary to adopt good research data management methodologies. Good data management is essential for both productive research and optimal use of the new data infrastructures. Effective management of research data is crucial for generating economic and scientific progress and for preserving this capital for future generations.

The activities related to most data infrastructures at the national and international level tend to focus on technical developments or technical standards for data management. Data management should also, however, focus explicitly on developing the social infrastructure needed to build knowledge and skills. In order to manage their own and others' data so that they can be re-used, researchers must have the necessary knowledge, skills and support available. Funders of research and research institutions can support and assist researchers in managing their data by requiring certain interventions or by installing tools and employing expert staff.

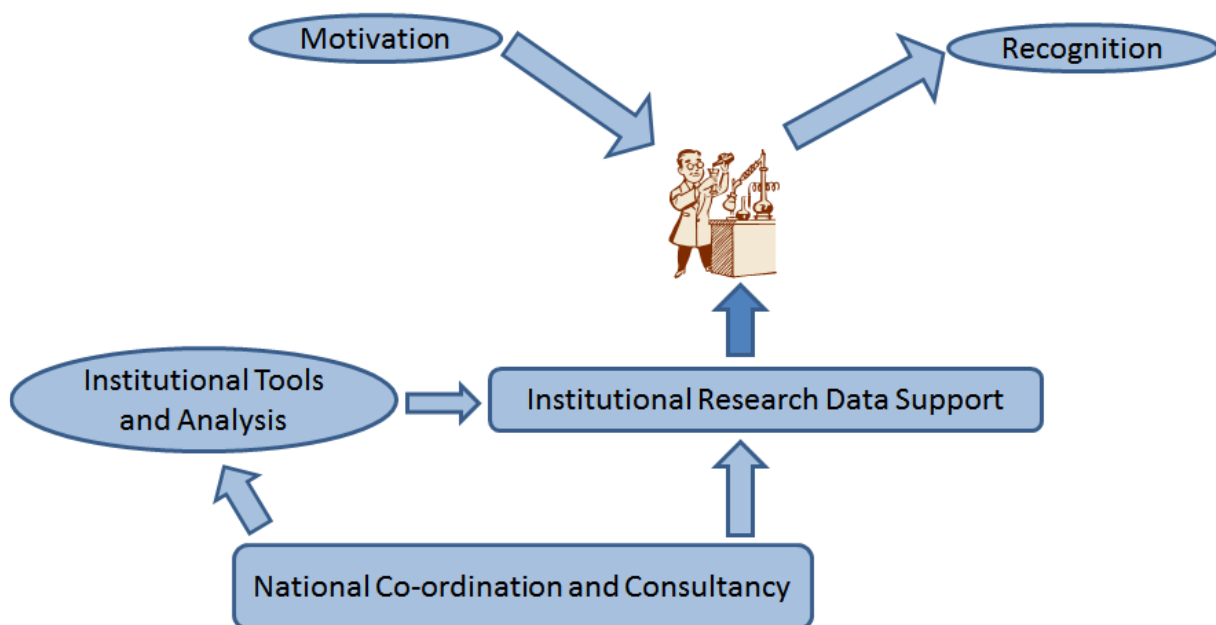
As yet it is unclear what interventions funding programmes, funding agencies or research institutions have available to ensure that researchers are supported in the management of their data. It is also not clear whether and to what extent such interventions are successful in achieving the relevant goals.

It is therefore desirable to produce an overview of possible interventions that have proved effective in supporting researchers in their data management. SIM4RDM contributes to this by analysing existing pan-European and international grant programmes and policies of institutions. It examines whether such programmes or policies include interventions that support researchers in gaining the knowledge, skills and support needed for data management. This report presents the results of an online survey to establish which interventions are already being used by funding agencies, research institutions, national bodies and publishers across the European Union member states and other countries outside Europe in order to improve the capacity and skills of researchers in making effective use of research data infrastructures. It also makes recommendations that organisations can adopt to help their research.

1 Methodology

The project team chose online questionnaires and supplementary desk research as instruments to determine what priority research funders, research institutions and other stakeholders in the data research environment give to interventions that enable researchers to improve their skills and ability to manage their research data. A workshop in which the survey results were augmented, was organised shortly after the survey had ended. Finally, the survey was enhanced with in-depth interviews.

The structure of the questionnaires is based on the target key areas identified by the Jisc Managing Research Data Programme model for developing social infrastructures. Representatives of the stakeholder groups were invited and provided additional information about possible recommendations for the elements of a framework for the subsequent areas: motivation, recognition, institutional tools and analysis, institutional research data support and national co-ordination and consultancy.



Arranging the questions according to the key areas of the Jisc Managing Research Data Programme model led to multiple questions in four categories, leaving out the key area of national co-ordination and consultancy. The survey consisted of questions about policies related to data management, recognition, institutional data support and institutional tools and analysis. Separate questionnaires were developed for national bodies. Examples of national bodies are the Digital Curation Centre¹ or Digital Archiving and Networking Services². These questionnaires consisted of seven (7) questions concerning their role in research data management.

¹ 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. See <http://www.dcc.ac.uk/about-us>

² www.dans.knaw.nl/en/content/about-dans

A growing number of reports and studies highlight the importance of good data management in order to avoid being overwhelmed by the data overload and to benefit from the use and re-use of data. Many of these studies offer short-lists of possible actions or recommendations to create technological or social data infrastructures. The suggested actions and recommendations were used as a basis for the questionnaires. The questionnaires were sent by e-mail or could be accessed via a website. There were different questionnaires for four (4) different stakeholder groups: research funders, national bodies, research institutions and publishers. Members of the Co-ordination Group of the Dutch Research Data Forum³ and the Research Data Management Working Group of Knowledge Exchange⁴ tested the questionnaires in advance.

The questionnaire is mainly geared towards answering questions dealing with research data management after a research project has been finished. This has been done to maintain a clear scope for this report and to avoid tackling a subject which is still hard to grasp. Input from the community, for example the SURF project Governing the Cloud⁵, indicated that research data management during a research project is still 'uncharted territory' which is often up to faculties or even research groups themselves to manage. Also, data management practices during a research project tend to be completely different when looking at various disciplines. So in order to avoid scope creep and researching badly defined subjects, this report will mainly focus on research data management at the end point of a research project.

The questionnaires were sent by e-mail to contacts of the project participants. Further distribution took place via mailing lists and publication on the website. European funders were identified based on an internet search and their membership of the European Science Foundation.

In order to draw valid conclusions, the project team had to receive at least 20 responses to the questionnaire. How many responses were received, can be located in the analysis of the policies of the respective bodies and organisations.

³ More information can be found at:

<http://www.surf.nl/en/actueel/Pages/Collaboratingonimprovedaccesstoresearchdata.aspx>

⁴ More information can be found at: <http://www.knowledge-exchange.info/Default.aspx?ID=285>

⁵ More information can be found at: <http://www.surf.nl/nl/projecten/Pages/RegieindeCloud.aspx>

2 Desk research

2.1 Introduction

Prior to creating the questionnaire for the various stakeholders involved in data management, a concise desk research was carried out in order to identify and analyse possible interventions for better data management. The main focus was on those elements that form part – or should form part – of policy and plans for data management by the various parties.

2.2 Riding the Wave

An important starting document was the '*Riding the Wave*' report that was published in October 2010⁶. This report reflects the vision of the High level Expert Group on Scientific Data on the infrastructure needed to manage scientific data in 2030. The report gives input to the European Commission for formulating policies for the research infrastructure within the framework of the Digital Agenda. '*Riding the Wave*' focuses on the infrastructure needed to manage scientific data. It identifies the benefits of accelerating the development of a fully functional e-infrastructure for scientific data⁷. The report continues with several actions to be undertaken by various EU institutions.

In the vision of the High Level Expert Group, data itself will become the infrastructure on which Europe can advance. The report gives examples of the growth of data and the increasing use within the distinctive disciplines and concludes that the way we do science will be changed totally.

To reach the vision of a collaborative data infrastructure that will enable access for researchers, use, re-use and reliability of data, the report makes six (6) recommendations⁸. Important for this project are the recommendations regarding developing and using new ways to measure data value, rewarding those who contribute to it, training a new generation of data scientists and broadening public understanding. When facing up the challenges, the report mentions a list of pitfalls, for example: the preservation of data, the protection of the integrity of the data, the conveyance of the context and provenance of the data and the protection of the privacy of the individuals to the data.

A follow up to the report of the High level Expert group was given by the four partners of Knowledge Exchange⁹ in November 2011. In the subsequent report '*A SURFboard for Riding the Wave*'¹⁰ Knowledge Exchange outlined a possible action programme. It identifies four key drives: (1) incentives, (2) training in relation to researchers in their role as data producers or users, (3) infrastructure and (4) funding the infrastructure in relation to further developments in data logistics. Researchers acting as data producers could be motivated to share and publish their data in four main areas: (1) re-use and recognition, (2) principles of science reflected in rules and codes of conducts, (3) requirements by funding organisations and (4) journal data availability policies. Regarding the training the report mentions that not only researchers should have basic skills with regard to data management but that new professionals as data librarians and data specialists are

⁶ *Riding the wave How Europe can gain from the rising tide of scientific data Final report of the High Level Expert Group on Scientific Data*, A submission to the European Commission, European Union, 2010

⁷ Idem 4, p.4

⁸ The six recommendations are: (1) develop an international framework for a Collaborative Data Infrastructure, (2) earmark additional funds for scientific e-infrastructure, (3) develop and use new ways to measure data value and reward those who contribute to it, (4) train a new generation of data scientists and broaden public understanding, (5) create incentives for green technologies in the data infrastructure, (6) establish a high-level, inter-ministerial group on a global level to plan for data infrastructure.

⁹ Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research More information available from <http://www.knowledge-exchange.info/>

¹⁰ Van der Graaf, M., L. Waaijers, *A SURFboard for riding the wave Towards a four country action programma on research data*, Knowledge Exchange, Denmark November 2011

needed. A data librarian is distinguished from a data specialist: the latter is part of a team of researchers or in close collaboration with them and the data librarian comes from the library community and is specialized in curation, preservation and archiving of data.

2.3 OECD Principles and Guidelines

Another set of fundamental documents are the '*Principles and Guidelines*' published by the OECD in 2007. These provide a number of recommendations for the sharing and reuse of research data using public funds for the purposes of producing publicly accessible knowledge.¹¹ The aim of the '*Principles and Guidelines*' is to increase the effectiveness and efficiency of science and scholarship; they are intended for research support and funding organisations, research institutions, and researchers themselves.

The '*Principles and Guidelines*' apply the following definition of research data: factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research and commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

The definition does not cover laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (for example laboratory samples, strains of bacteria and test animals such as mice). Access to all of these products or outcomes of research is governed by different considerations than those dealt with here.¹² The '*Principles and Guidelines*' concern access arrangements: arrangements between research institutions, research funding agencies, and other parties to determine the conditions for access to and use of data.

The Principles drawn up by the OECD have been taken into account when drawing up the questionnaires; agreements on access to and use of research data are a component of data management policy and plans but these comprise more than regulating access to data.

2.4 Data Management Policies

A number of guidelines have appeared in recent years for policy-making regarding research data. Finland strives for a clear national data policy by establishing legislation, responsibilities, roles, operating models and payment policy.¹³ Despite the aim of this policy transcending the objectives of SIM4RDM, various elements can be distilled from these guidelines. The important factors include the recommendations to determine the principles for the availability of data and to develop competence in data production and management, including quality training, data analysis, data management, metadata work and competence in infrastructure.

Policies in the United States

'*Circular A-110*' of the Office of Management and Budget sets forth standards for obtaining consistency and uniformity among federal agencies in the administration of grants to and agreements with institutions of higher education, hospitals, and other non-profit organisations¹⁴. A number of US funding agencies have drawn up data management policies based on that circular.

¹¹ *OECD Principles and Guidelines for Access to Research Data from Public Funding*, OECD 2007

¹² *Idem* 1, p.13

¹³ *Research Data – Guide for Policy-makers*, available from: http://www.sim4rdm.eu/sites/default/files/Research%20Data-Guide%20for%20Policy%20Makers_0.pdf

¹⁴ Office of Management and Budget, *Circular A-110 REVISED 11/19/93 As Further Amended 9/30/99*, Available from http://www.whitehouse.gov/omb/circulars_a110/

Since 2011, the National Science Foundation in the United States has made a data management plan mandatory when submitting grants proposals¹⁵. Proposals must now include a supplementary document of no more than two pages labelled 'Data Management Plan' which should include the following information:

- **Products of the Research:** The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.
- **Data Formats:** The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate this should be documented along with any proposed solutions or remedies).
- **Access to Data, Data Sharing Practices and Policies:** Policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.
- **Policies for Re-Use, Re-Distribution and Production of Derivatives**
- **Archiving of Data:** Plans for archiving data, samples, other research products, and for preservation of access to them.

The National Aeronautics and Space Administration (NASA) considers data to include observation data, metadata, products, information, algorithms, including scientific source code, documentation, models, images, and research results. NASA's policy involves a commitment to the full and open sharing of Earth science data obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users as soon as such data become available. There is no question of exclusiveness: following a post-launch checkout period, all data will be made available to the user community¹⁶.

The National Oceanographic Data Center (NODC)¹⁷ is the United States facility established to acquire, process, store, and disseminate oceanographic data from the United States and other countries. Its '*Long Version of the Data Submission Guidelines*'¹⁸ mentions that the Federal Ocean Data Policy requires that appropriate ocean data and related information collected under federal sponsorship to be submitted to and archived by designated national data centres.

The National Institutes of Health (NIH) consider all data eligible for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data. To facilitate this, investigators submitting a research application to NIH requesting USD 500,000 or more of direct costs in any single year, are expected to include a plan for sharing final research data for research purposes or state why data sharing is not possible¹⁹.

Policies in the UK

In January 2007 a snapshot was published of the then current policies and practices of major UK research funders²⁰. Although policy at the time of the study focused primarily on access to journal articles and conference proceedings, it was recognised that it is important to devote attention to research data. It was found that faulty coordination of policies for collecting and managing data sets could mean the loss of important data. It was also noted that some research councils had a pretty good infrastructure with associated policy on data curation.

¹⁵ National Science Foundation, *Proposal and Award Policies and Procedures Guide*, 1 January 2011, available from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#IIC2j

¹⁶ National Aeronautics and Space Administration, *Data and Information Policy*, available from <http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/>

¹⁷ Available from <http://www.nodc.noaa.gov/General/NODC-Submit/submit-guide.html#polguide>

¹⁸ National Oceanographic Data Center, *Long version of the Data Submission Guidelines*, available from <http://www.nodc.noaa.gov/General/NODC-Submit/submit-guide.html>

¹⁹ National Institutes of Health, *Data Sharing and Implementation Guidance*, 5 March 2003, available from http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

²⁰ *Research Funders' Policies for the Management of Information Outputs*, A report commissioned by the Research Information Network January 2007 available from <http://rinarchive.jisc-collections.ac.uk/our-work/research-funding-policy-and-guidance/research-funders-policies-management-information-output>

At the same time, there were major differences in the extent to which research funding bodies see it as their responsibility to provide long-term data storage and accessibility. It was observed that the differences in the nature and origins of data bring with them differences in value, and implications for policy and practice.

The snapshot also showed that universities allocate responsibility for data management to researchers themselves or to their departments. They saw data curation as a specialist task for which they did not consider themselves appropriate. This is why they do not encourage the depositing of data in repositories.

The study advised funders to develop policy keyed to this kind of research, the institutional and funding environment, and the broader policy imperatives of funders and research institutions. It called for broader cooperation in a number of fields, including the division of roles and responsibilities, and the preservation and curation of valuable research results. Data and information policies are extremely important in this context. Cooperation is necessary between funders, research institutes, and specialised agencies.

The year 2008 saw the publication of *Stewardship of Digital Research Data: A Framework of Principles and Guidelines*²¹. That document addresses some of the key issues that arise in managing the unprecedented quantities and varieties of digital data now being created and collected by researchers. It sets out a policy framework of five principles, with associated guidelines, to help ensure that such data are properly looked after. The five principles are as follows:

1. The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with the principles set out in this document.
2. Digital research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others should include proper quality assurance.
3. Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data is made accessible and used.
4. The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.
5. Digital research data of long-term value arising from current and future research should be preserved and remain accessible for current and future generations.

The report *Dealing with Data*²² published in 2007 investigates the various roles and responsibilities of parties in the United Kingdom that deal with data. The report offers a number of recommendations to the Joint Information System's Committee (Jisc), funding agencies, and institutions in a number of different categories. Research funding bodies are advised to publish, implement and enforce a data management, preservation and sharing policy. Submission of a structured data management plan should be an integral component of funding applications. Not only research funders, however, but also every higher education institution should implement a data management plan with a recommendation to deposit data in an appropriate open access data repository and/or existing data centre. All relevant stakeholders are advised to identify incentives to encourage researchers to store their data in an appropriate open access data repository.

²¹ *Stewardship of Digital Research Data: A Framework of Principles and Guidelines. Responsibilities of research institutions and funders, data managers, learned societies and publishers*, Research Information Network, January 2008, <http://rin.archive.jisc-collections.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf>

²² Lyon, L., *Dealing with Data: Roles, Rights, Responsibilities and Relationships*

In 2012, Research Councils UK (RCUK) provided seven common principles on data policies which provide an overarching framework for individual Research Council policies on data²³. In accordance with the Principles, institutional policy and data management plans must be in line with relevant standards and community best practices. Where discoverability and reuse are concerned, sufficient metadata should be included that are openly available to others.

An indication should also be given of how the underlying data can be accessed. Throughout all phases of the research process, account must be taken of legal, ethical and commercial restrictions on the release of research data. The RCUK provides scope for a limited period of privileged use before the release of data: the length of this period depends on the particular research discipline. Users of data must indicate the source and must observe the conditions set regarding access to the data.

The Digital Curation Centre (DCC) has published an overview of the data policy of the individual British research funders²⁴. A table clearly indicates the elements that they have included in their policy and the support that they provide. In this table the DCC defines 'data' as 'a datasets policy or statement on access to and maintenance of electronic output'. The elements distinguished by the DCC also include set timeframes for making content accessible or preserving outputs, a data plan, access and sharing, long-term curation, monitoring whether compliance and possible action have been taken, the provision of a data centre to curate unpublished electronic resources or data, and a willingness to meet publication fees and data management/sharing costs.

The Data Audit Framework²⁵ was developed by Jisc in 2007 in the light of the '*Dealing with Data*' report as a framework that institutions can use to collect information regarding what data they hold and how it is managed. This tool enables institutions to improve their activities in the area of data management²⁶. Various pilots to test the tool have revealed that researchers require elementary training and guidance as regards creating and managing their digital assets. No differences could be identified between disciplines or institutions regarding such matters as loss of data, irretrievability, and lack of storage. It was concluded in the light of the pilots that drawing up an institutional data policy would be a valuable initial step towards tackling the problem of data management. It became apparent that only a fraction of the data is managed by specialised data centres and that long-term preservation is often the concern of individual departments, which have hardly any of the necessary skills or capacity.

2.5 Needs of researchers

Whether data can be reused depends entirely on good data management. The use of data for purposes that were not known in advance perhaps creates more challenges regarding management of the data. Cooperation by researchers is therefore very important. The '*PARSE*' report²⁷ published in 2009 shows that researchers are not enthusiastic about sharing their data. The most frequent reasons given for this are unfamiliarity with possible data storage locations, fear of the data being misused or interpreted wrongly, and legal issues. The authors of the '*PARSE*' report make a number of recommendations regarding these matters. Encouraging researchers to share their data and rewarding them for doing so would ensure that they do so to a greater extent. The authors also recommend defining standards for the openness of data and applying those standards to the exchange of data sets between research institutions and repositories. Alerting researchers to the possibility of linking and citing data within and beyond disciplines would also encourage them to share their data.

²³ *Excellence with Impact: RCUK Common Principles on Data Policy* at <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

²⁴ Overview of funders' data policies at <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

²⁵ <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2007/dataauditframework.aspx>

²⁶ Sarah Jones, *The Data Audit Framework: A First Step in the Data Management Challenge*, *International Journal of Digital Curation*, issue 2, Volume 3, 2008

²⁷ PARSE.Insight *Deliverable 3.6, Insight into digital preservation of research output in Europe*, p.46 available from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf

One way of ensuring researchers' cooperation would be to provide support for managing their data, with proper account being taken of their wishes and requirements. SURF has carried out a survey of what those wishes and requirements actually are²⁸. The survey showed that researchers feel a need for support regarding data storage because they lack the skills, knowledge, and awareness necessary to improve their day-to-day storage of data.

The support provided should comply, however, with a number of requirements. Researchers do not want measures to be imposed 'from on high' and/or made obligatory. They want to be in control of what is done with their data, of whom they share it with, and of the conditions for this to be done. That means that they must be able to rely on the party that manages their data – a data centre, library, etc. – respecting their interests.

Where the aids and services provided are concerned, researchers ask that these be user-friendly and keyed to their workflows, which often differ according to the particular discipline (and sometimes even according to the project). They expect that the tools and services support them in their day-to-day work within the research project; the long-term or general interest should be subordinate to that purpose. Finally, the advantages of the support should be immediately apparent, with the support being provided locally, being practical, and being available when it is needed.

The outcome of the survey is confirmed in actual practice. In the CARDS project²⁹, researchers were assisted in managing their data in the most efficient and effective manner. This made clear that it is important to formulate a data management policy, to provide effective support, and to construct and maintain an effective infrastructure for data storage and data management.

The Incremental report³⁰ published in July 2010 finds that the problems that researchers have with data management concern simple day-to-day matters such as the lack of a file management system or a naming system, making it unclear what format can best be utilised in the medium term. Researchers also face ad hoc problems regarding data storage. They are also averse to jargon; the language used needs to be clear and technical terms need to be avoided. This helps not only when addressing the problems and risks that researchers experience but also when providing data management services. Researchers would seem to be confused by such terms as 'digital curation', and most of them do not know what a 'digital repository' actually is. Many people are suspicious of 'policies', which sound like a hollow mandate, but are receptive to 'procedures' or 'advice' which may be essentially the same thing, but convey a sense of purpose and assistance rather than requirement³¹.

The conclusion is that the best point at which to intervene is at the start of the researcher's career, when he or she can be provided with guidance, training, or management tools. The training of PhD students and post-doctoral researchers can also be valuable because at this stage they have yet to adopt the habits and practices of more senior colleagues.

There is appreciation in various fields for the need to preserve and reuse data. The US National Academy of Sciences has made a proposal to ensure access to data for future generations³². The Academy notes that data management and data preservation generally enjoy only low priority.

²⁸ Feijen, M. *What researchers want. A literature study of researchers' requirements with respect to storage and access to research data*. SURFfoundation Utrecht February 2011 available from http://www.surf.nl/nl/publicaties/documents/what_researchers_want.pdf.

²⁹ Heesakkers, D, A. van Meegen, *CARDS Controlled Access to Research Data, Stored Securely* SURF Utrecht January 2012

³⁰ Freiman, L., C. Ward, S. Jones, L. Molloy, K. Snow, *Incremental. Scoping study and implementation plan*. A pilot project for supporting research data management, University of Cambridge, University of Glasgow, July 2010 available from: http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf

³¹ idem 6, p. <>

³² National Academy of Sciences, *Preserving Scientific Data on our Physical Universe: A New Strategy for Archiving the Nation's Scientific Information Resources*, 1995

The report identifies a number of problems in observational databases: there are significant shortcomings as regards documentation, access, and the long-term preservation of data in usable form. There is also a lack of directories that specify what data sets exist, where they are stored, and how they can be accessed. The challenge is to develop data management and archiving procedures that can cope with the rapid increase in the volume of data but that are also able to access existing data sets³³.

More recently, a number of organisations have also studied data management practices and requirements amongst scientists in different fields. One example is EUDAT, a European organisation which 'aims to address these challenges and exploit new opportunities using its vision of a Collaborative Data Infrastructure'³⁴. In its 'Data Management Landscape Characteristics and Community Requirements' report, published in April 2012, they focussed on the community requirements and on identifying common services suitable for a collaborative data infrastructure. They did this by conducting interview with a number of experts from a number of important data communities, such as CLARIN, LIFEWATCH and VPH. One important conclusion of this report is that within the core communities involved, there 'is a clear need for safe and dynamic data replication services, in particular the safe replication service'³⁵. Furthermore, they identified a number of other relevant services, including EUDAT Metadata Domain, Researcher Data Store and a Persistent Identifier service.

A similar study has also been conducted by EarthCube, a US community aimed at integrating the cyberinfrastructure of the earth sciences in the US. In their roadmap to cross-domain interoperability, they concluded that sound data management requires 'support for dataset documentation and curation from the beginning of its life cycle; development of discovery mechanisms that operate in a federated system of catalogs with different domain contexts; tools to support data exploration and manipulation to extract the desired information; a social framework for cross-domain networking between researchers needing to understand each other's data; and a governance system to provide direction, decision-making, and authority for prioritizing and developing the necessary specifications and tools'³⁶.

In a research project very similar to SIM4RDM, the US DataONE community conducted a large-scale survey of data sharing practices and data management plans of scientists, academic libraries and librarians, and data managers. In their study of scientists, conducted among more than 1300 scientists from both the US and international partners and from different disciplines, they concluded that the majority of the respondents were willing to share their data and re-use other data if there were some restrictions and conditions to use and re-use³⁷. Other important conclusions are that scientists do not believe their institutions are sufficiently helping them in long-term data preservations and that there is a lack of awareness of the importance of meta-data. Lack of funding and insufficient time were reasons scientists did not make their data electronically available, even though they indicated that their ability to answer scientific questions was restricted, because they could not access the data needed³⁸. The same questions were also asked to data managers and they indicated the major reasons for not making the data electronically available were lack of incentives and the P.I. not wanting to share the data. Conditions for sharing the data mostly referred to formal acknowledgement and/or citation, and a clear overview of who is using their data³⁹.

³³ idem 11, p.

³⁴ <http://eudat.eu/about>

³⁵ EUDAT, *Data Management Landscape Characteristics and Community Requirements*, April 2012, p. 28, available from <http://eudat.eu/deliverables/d411-data-management-landscape-characteristics-and-community-requirements>

³⁶ EarthCube Roadmap, Cross-Domain Interoperability Test Bed Group of EarthCube, august 2012, p2, available from https://www.dropbox.com/s/0oqk5ostahfokbg/interop_roadmap_master8_Aug16.pdf

³⁷ Tenopir, C. et al. Data Sharing by Scientists: Practices and Perceptions, p. 18, available from <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0021101>

³⁸ Tenopir, C. et al. Data Sharing by Scientists: Practices and Perceptions, p. 20, available from <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0021101>

³⁹ Read, E. J., Birch, B., Tenopir, C., Frame, M., & Zolly, L. (2012). DataOne: A glimpse into the practices of data managers. A presentation given at the 38th Annual International Association for Social Science Information Services and Technology (IASSIST) Conference, Washington, D. C.

2.6 Data Management Plans

The above-mentioned study by the National Academy of Sciences and DataONE make it possible to identify guiding principles for drawing up data management plans. Data is the lifeblood of science and its value is in its reuse. Access to data is therefore just as important as acquiring and preserving it. These elements already need to be taken into account in the initial stage when creating the data.

A very clear summary of the content of a data management plan is given in an Australian guide to data management published in 2008⁴⁰. This stipulates that a data management policy should deal with the following: the research discipline of the project; how the research is to be conducted; the funding arrangements for the research project; the kind of data generated or collected by the project; how and when data is to be deposited into a database or repository; when and on what basis data is to be shared and made available for access by other researchers; any legal obligations imposed on the research project or individual researchers; and how intellectual property rights are to be managed⁴¹.

The study also looks closely at data management plans. The writers apply the definition formulated by the Australian National University in September 2010: a document that describes what data will be created during a project and how it will be managed. In particular it is a document that describes what research data will be created, what policies (funding, institutional, and legal) apply to the data, who will own and have access to the data, what data management practices (backups, access control, archiving) will be used, what facilities and equipment will be required (hard-disk space, backup server, repository), and who will be responsible for each aspect of the plan.⁴²

Refining the definition produces a number of topics that should be included in every data management plan: data ownership and responsibilities, legal rights, data security and sustainability, access and re-use. Where reuse is concerned, further details should include: which data is to be made accessible, when the data is to be made accessible, who may access the data, how the data will be made accessible, how widely the access rights should be granted and metadata⁴³.

In order to comply with the requirements set by the National Science Foundation regarding the provision of funding, Rice University has made recommendations for the content of data management plans. According to those recommendations, a data management plan should:⁴⁴

- be as simple as possible while still doing what is necessary
- describe the data
- present the context of the data
- explain the nature of the data
- describe the method for preserving and/or curating the data
- discuss the approach for accessing the data, if relevant
- state how long the data will be preserved and/or curated
- clarify ethical and/or privacy issues associated with the data, if relevant
- detail intellectual property concerns associated with the data, if relevant

⁴⁰ Fitzgerald, A, K. Pappalardo, A. Austin, *Practical Data Management: A legal and policy guide*. September 2008, version 1.0 available from: http://eprints.qut.edu.au/14923/1/Microsoft_Word_-_Practical_Data_Management_-_A_Legal_and_Policy_Guide_doc.pdf

⁴¹ *Idem* 11, p.29

⁴² The Australian national University, *ANU data management manual, Literacy Program, September 2010, p. 5*, available from http://regnet.anu.edu.au/sites/default/files/files/ANU_Data_Management_Manual.pdf

⁴³ *Idem* 11. p.41

⁴⁴ <http://osr.rice.edu/forms/dataManagementPlans.pdf>

The British document '*Policy-making for Research Data in Repositories: A Guide*⁴⁵ published in 2009 can help institutions to decide on and plan data management. The guide comprises requirements and examples for the various topics that can help draw up a data management plan and that can be included in it. It contains a set of data-related topics that focus on data quality, management, and preservation. The chapters describe content coverage, various types of metadata, the submission of data (ingest), access and reuse of data, preservation of data, and withdrawal of data and succession plans.

⁴⁵ Green, A, S. Macdonald, R. Rice, *Policy-making for Research Data in Repositories: A Guide*, version 1.2, May 2009 available from: <http://www.disc-uk.org/docs/guide.pdf>

3 Analysis of the policies

3.1 Research Funders

The online survey consisted of seven (7) questions regarding the funding organisation, its tasks, funding and its perceived maturity. In total, eighteen (18) research funders participated in the survey. The geographical distribution of the respondents is shown in table 1.

COUNTRY	RESPONDENTS
FINLAND	3
NORWAY	3
IRELAND	2
SWEDEN	2
BULGARIA	1
HUNGARY	1
UK	1

Table 1: Research funders

3.1.1 Policy for funding and data management

Nearly half of the organisations have a funding policy covering data management. Of the organisations without a policy covering data management, 35% have no intention of having one, 10% think it will take more than 24 months before their policy will cover data management, and 5% think this will take 6 to 11 months.

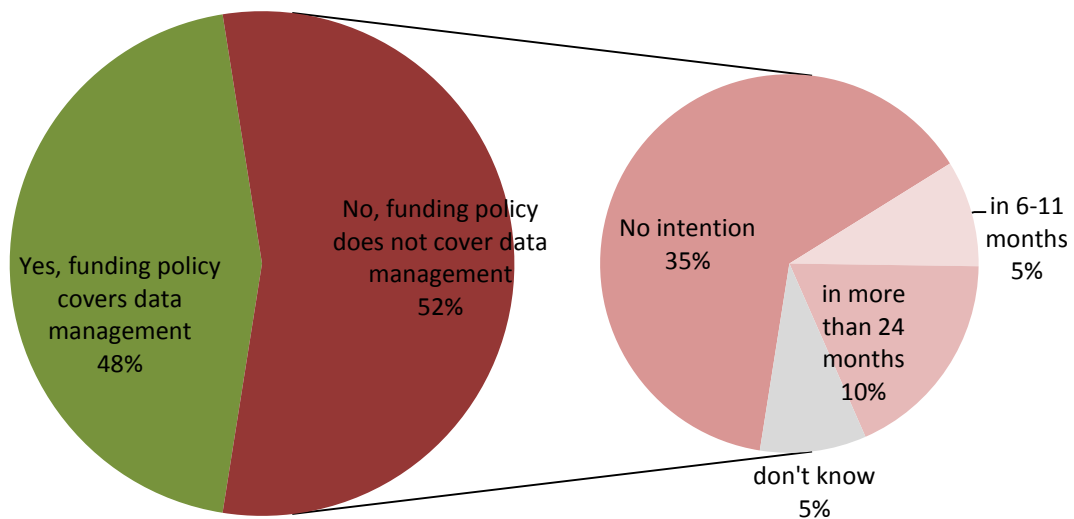


Figure 1: Policy for data funding and data management (n=18)

3.1.2 Requirement of Data Management Plan

Of the research funders, 25% require a Data Management Plan as part of the grant application, 10% recommend it and 65% do not require a Data Management Plan.

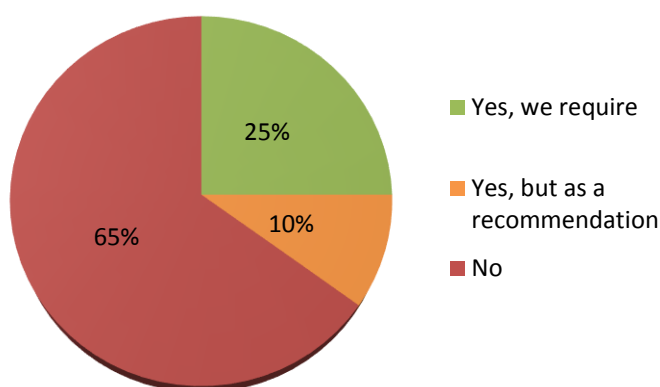


Figure 2: Data Management Plan (n=18)

Drivers for requiring a Data Management Plan vary and range from initiatives supporting open data, and better research and science to initiatives by individuals who have co-operated with others in other organisations. Respondents mentioned that publicly funded research must be open. Other drivers are the increase in the amount of data and the fact that data-intensive research is more and more common. Furthermore, research funders see the need to curate and make key datasets created with their funding discoverable, and they want to see links to grant identifiers and research outputs. The main reason Enterprise Ireland⁴⁶, the government organisation responsible for the development and growth of Irish enterprises in the world's markets requires a Data Management Plan, is to drive research and commercialisation. Research projects must be fully documented to apply for patents, to license technology, or to engage effectively with companies.

Of those organisations that require or recommend a Data Management Plan, 83% do so for all disciplines.

Those organisations that do not require a Data Management Plan state several reasons for their policy. One of the funders said that this requirement had not been fully explored yet. For some funders, the main concern so far has been to disseminate results and extend knowledge and innovations to the end-users. Although a Data Management Plan is not required as part of a grant application for funding, if a proposal is successful, all the related research publications must be placed in open access repositories. Another reason for the absence of a requirement is that the area the funder covers is too heterogeneous (from music to medical research) for detailed recommendations. Teagasc⁴⁷ the Irish agriculture and food development authority, responded that most of the research it funds is conducted on its premises. Most of the data arising from the experiments are captured and stored in a secure database. As a result, this funder has not encountered any issues or difficulties with regard to data management, and so they do not see why they would need to put formal requirements in place. Finally, ownership of data created within institutional work was said to be problematic.

The respondents were asked whether their organisation requires certain elements to be addressed in the Data Management Plan. The results are shown in Figure 2.

⁴⁶ <http://www.enterprise-ireland.com/en/About-US/>

⁴⁷ <http://www.teagasc.ie>

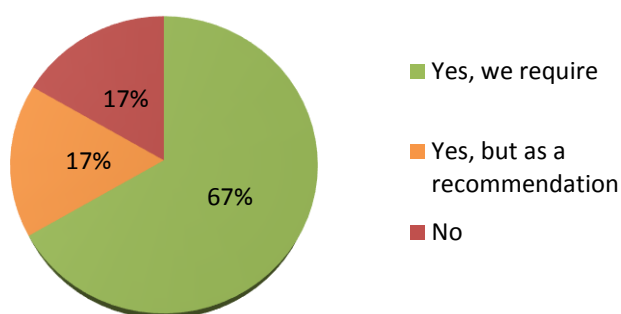


Figure 3: Does the organisation require certain elements in the Data Management Plan? (n=6)

What the survey showed is that several elements need to be addressed in the Data Management Plan. One of the funding agencies made the elements quite explicit: it describes how the research material will be obtained, how it will be used, how it will be stored and protected, how its subsequent use will be facilitated, and the rights of ownership and use.

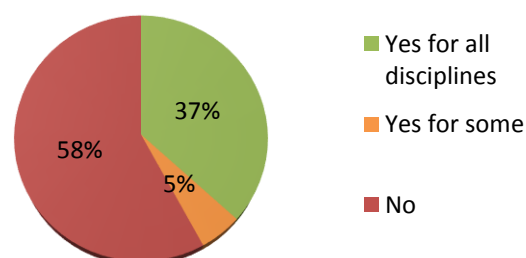
Enterprise Ireland breaks all research projects into work packages. Each work package must be documented on completion and supplied to Enterprise Ireland. Regular site visits are carried out to validate technical and commercial progress.

One funder requires outline Data Management Plans for all proposals from July 2012. This organisation describes in detail how to deliver the data.⁴⁸This funder also requires a full Data Management Plan for successful proposals. Another research funder stipulates the type of metadata that will be tied to the data and when, where and how data is made available and who the contact person is.

Organisations gave no specific reasons for requiring certain elements in the Data Management Plan.

3.1.3 Data preservation

Just over one third (36%) of the organisations requiring retention of data designated a specific organisation for preservation. Several of the respondents mentioned these organisations by name⁴⁹.



⁴⁸ (1) A table with a row for each dataset expected to be produced and three columns headed 'Dataset Description', 'Release Date to Data Centre', 'Re-use Scenarios'. (2) An indication of which existing datasets will be used by the research project. (3) A description of the data management methods and descriptive metadata to be associated with each dataset before it is transferred to the NERC data centre.

⁴⁹ (1) Environment Climate Data Sweden (ECDS), (2) Svensk Nationell Datatjänst hosted by Göteborg University for Social Sciences, Humanities and Medicine, (3) Swedish LifeWatch for biodiversity data, (4) BILS (Elixir) Bio informatic, (5) BBMRI.se (Biobanks and related data), (6) Swedish ICOS (Carbon flux data), NERC (six wholly owned Environmental Data Centres and a 7th (science-based archaeology) which the research funder supports), (7) Norwegian Social Science Data Services (NSD)

Figure 4: Data retention (n=18)

3.1.4 Reservation of funds for support and storage of research data

Only one organisation (5%) requires a certain amount of the money to be reserved for support and storage of research data.

3.1.5 Length of time that data is required to be made accessible

Of the respondents, three (3) organisations (15%) stated how long data must be made accessible. For one this is 24 months, while for the others this is 84 months. One funder did not answer this question.

3.1.6 Length of time that data is required to be preserved

Figure 5 shows how long data must be preserved.

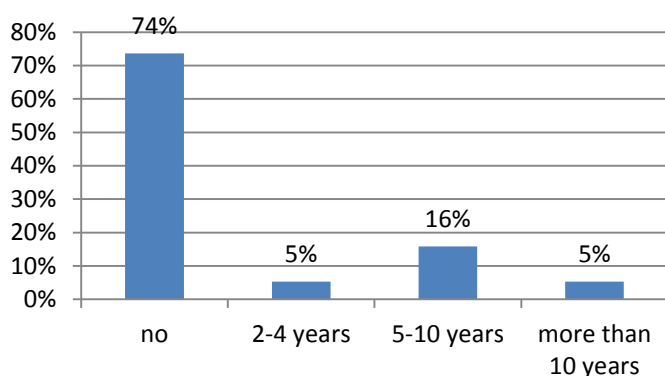


Figure 5: Term for how long data should be preserved

Several factors contribute to determine how long data must be preserved. Preservation for more than five years is determined by the function of the institute in question: the respondent is a museum that functions as a data archive. The data is considered an artefact and the main goal is to preserve data as long as possible. Another factor is the commercial use of data. A funder remarked that generally, commercialisation occurs within the first three years of project completion for industrial technology and ICT projects. Biotechnology projects take five or more years. In other cases, the data retention period is driven by audit requirements or the contract.

3.1.7 Evaluation of the Data Management Plan

Of the twenty (20) organisations, three (3) (15%) require reports submitted by funded projects to include an evaluation of the Data Management Plan. Legal issues, access policies and long-term archiving are elements that must be evaluated. The Natural Environment Research Council (NERC) based in the UK describes how a Data Management Plan should be evaluated⁵⁰. If data are not delivered satisfactorily to the data centre, sanctions are possible.

Some of the organisations that do not require an evaluation of the Data Management Plan defended their position by stating that evaluation of the Data Management Plan is part of an overall evaluation and should not be mentioned separately.

⁵⁰ All projects submit an Outline DMP. If a project is successful the Data Centre will take the ODMP and the research description ('Case for Support') and produce the first draft of the DMP. The Principal Investigator (PI) will amend this and 'sign-off' with the relevant Data Centre. If at the end of the grant datasets are not delivered satisfactorily to the Data Centre, then the PI and the Department may be open to sanctions (e.g. barring an investigator from making further applications to NERC).

As reported above, most of the research funded by Teagasc is conducted on its premises. Its method of capturing and securing data produced by its own research has allowed it to avoid any issues or difficulties with regard to data management.

Some funders have never raised the evaluation of a Data Management Plan as an issue. It was mentioned that this is still under development and that an evaluation will probably be carried out in the future. Data management requirements are currently being developed.

3.1.8 Experience

Four of the 20 organisations (20%) employ a mechanism to evaluate whether the requirements for the management plan have been fulfilled. They ask reviewers to ensure that the outline Data Management Plan is satisfactory. Reviewers are also asked to comment on and make recommendations about the plans.

3.1.9 Requirement that best advances research data management

The respondents were asked which requirement has worked best to advance research data management. The suggestion of storing research data with the data organisations available in the social sciences and linguistics worked well. Another requirement that worked well was to define clear deliverables for the research projects.

Other requirements that were mentioned were detailed instructions for writing the research plan, the implementation of a central laboratory information management system and data archiving. One suggestion was to ask grant applicants to state how they will deal with data management and data openness and then later hold them accountable for their words.

3.1.10 Conclusion

Nearly half of the research funding organisations that responded to the survey have a policy covering data management. A quarter of those require a Data Management Plan as part of the grant application, in most cases for all disciplines.

The most important drivers for requiring a Data Management Plan are better research and science, and support for open data and access. The increase in the amount of data and the fact that data-intensive research is more and more common are other reasons. Funders see the need to curate and make data sets created with their funding discoverable. Commercialisation is another motive: research projects must be fully documented in order to commercialise their results effectively.

Only 15% of the research funders require an evaluation of the Data Management Plan. Respondents indicated that data management is still under development. The plans will probably be evaluated in the future.

Of the funding organisations without a policy, 35% have no intention of establishing one. For some funders the main concerns so far have been to disseminate the results and to extend knowledge to end-users. The area covered by the funding was also too heterogeneous, or the research was conducted on the premises so that formal requirements were not needed.

Ownership of data created within institutional work was mentioned as a problematic issue.

A Data Management Plan should address how data is obtained, used, re-used, stored and protected. The rights of ownership and use must also be part of the plan. Regarding the question of when data is to be made accessible, one research funder stated that data must be re-used within 24 months; for the others, this is seven years (84 months).

Just over one third of the organisations designate a specific organisation for preservation and mentioned national organisations by name. Only 25% of the funders stated how long the data must be preserved.

3.2 National bodies

Institutions with a contact person known to the project participants received an invitation by e-mail to participate in the survey. The invitation was also distributed to mailing lists. The online survey consisted of seven questions regarding the national body, its tasks, funding and its perceived maturity. A total of eighteen (18) national bodies participated in the survey, their geographical distribution is shown in table 2.

COUNTRY	RESPONDENTS
FINLAND	8
THE NETHERLANDS	3
UK	3
AUSTRALIA	1
HELSINKI	1

Table 2: National bodies

3.2.1 National body for data management

Over one third (38%) of respondents reported that their country has a national body co-ordinating activities related to research data management.

None of the countries without such a national body is planning to set up one.

3.2.2 Funding of the national body

Figure 6 shows the funding of national bodies. Other sources of funding mentioned were private companies (2x) and the Joint Information Systems Committee (Jisc).

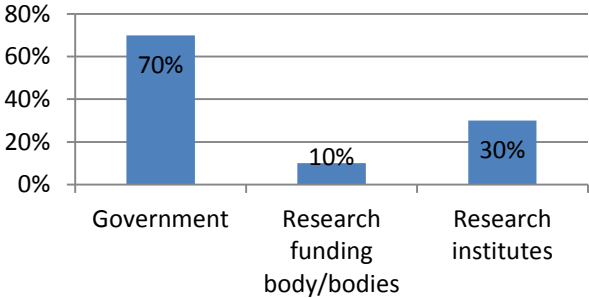


Figure 6: Funding of national bodies (n=10)

3.2.3 Maturity of the national body

Those respondents reporting a national body were asked about the maturity of this organisation. The answers are shown in Figure 7.

Nearly all national bodies have the status of a project. The organisation that is a self-sustained legal entity still considers itself a project organisation regarding research data.

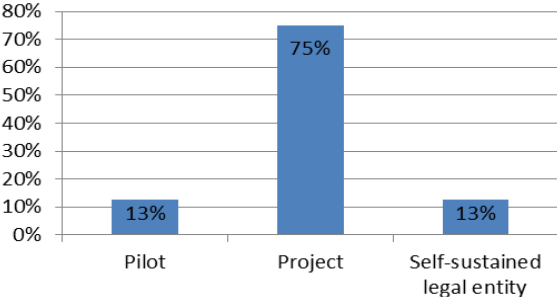


Figure 7: Maturity of national bodies (n=8)

3.2.4 Inspiration from across the border

All of the national bodies looked across the border for inspiration. Asked what they had taken away from these examples, they gave various answers. National bodies looked at the series of reports by Jisc⁵¹/UKOLN⁵²/RIN⁵³ to identify the functions that a national entity should provide. They also mentioned best practices, (semi-)mature solutions, and various cultural aspects. Working closely, with partners in the EU and the USA, including in joint infrastructure projects, in order to learn from what others have or haven't done was said to be immensely useful for informing practice. It was brought to the attention of the project team that a number of excellent initiatives are currently under way to produce tools, resources, support and guidance for improved research data management. The vast majority of these initiatives are funded through short-term projects, however, most of them running for just one to three years. The lack of long-term funding at the national level continues to pose a challenge to the sustainability of these initiatives.

3.2.5 Tasks of the national body

The respondents who said that their country had a national body were asked to state its tasks. Figure 8 shows the answers.

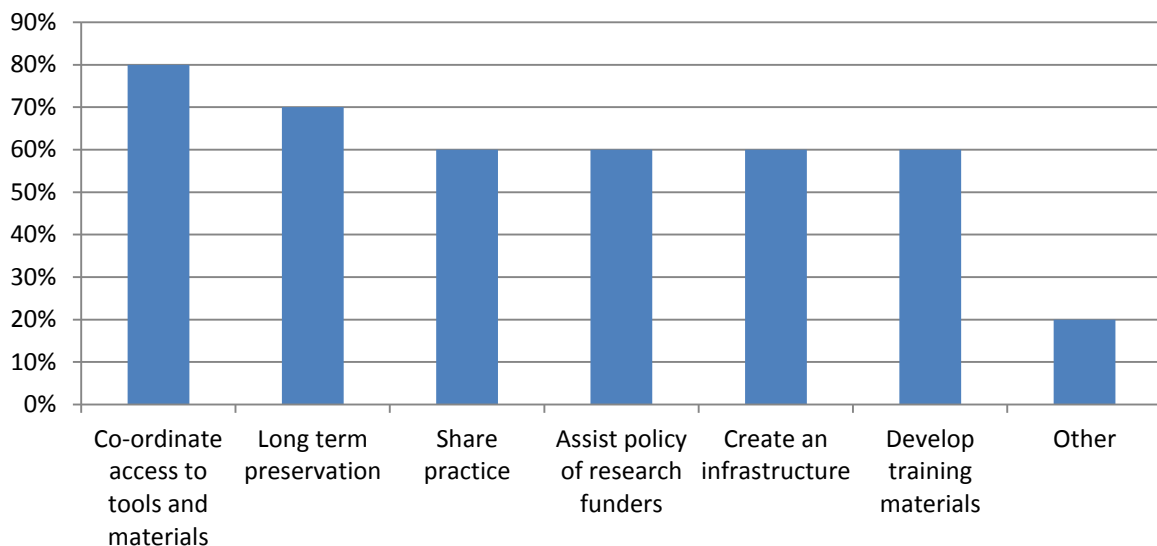


Figure 8: Tasks of national bodies (n=10)

National bodies also provide co-ordination and raise awareness. They fund data capture projects at institutions (software to better capture data and metadata) as well as projects that bring selected data under management retrospectively and make them discoverable. They also fund support for the creation of a national set of reference data collections.

⁵¹ More information about Jisc at <http://www.Jisc.ac.uk/aboutus.aspx>

⁵² More information about UKOLN at <http://www.ukoln.ac.uk/>

⁵³ More information about RIN at <http://rinarchive.Jisc-collections.ac.uk/>

3.2.6 Conclusions

Nearly 40% of the respondents reported that their country has a national body for coordinating research data management activities. In 70% of the cases, government finances these bodies. Other sources of funding mentioned are research institutes and, the research funders of private companies. Most of the national bodies are project organisations; only one is a self-sustained entity, although it stated that it is still a project organisation as regards research data.

The most important tasks of the national bodies are to coordinate access to tools and materials, long-term preservation to share practices, to support the research funders' policies and to create infrastructures.

3.3 Institutional policies

At the time that the survey concluded, 94 institutions had responded. The majority of the respondents originated from the project partner countries (see table 3).

The questionnaire regarding institutional data management policies is broken down into four categories: Policies and Data Management, Recognition, Institutional Research Data Support and Institutional Tools and Analysis. The category Policies and Data Management consists of four questions to determine what motivates a research institute to establish a policy. The first question aimed to discover whether an institution has or intends to create a policy. The second question regarded the year the policy was created or the timeframe in which an institution was about to generate a policy. The third question looked at the drivers for establishing a data management policy: this question explored whether funders can influence research institutions or whether institutions are driven by other developments. The last question examined the elements of existing policies. These elements were based on the recommendations found in the literature.

COUNTRY	RESPONDENTS
FINLAND	21
UK	16
SWEDEN	15
THE NETHERLANDS	11
IRELAND	8
DENMARK	6
NORWAY	4
HUNGARY	3
SWITZERLAND	3
AUSTRIA	2
AUSTRALIA	1
BELGIUM	1
LITHUANIA	1
PAKISTAN	1
USA	1

Table 3: Institutions

3.3.1 Policies and data management

Policy on data management

Over one third (37%) of the institutions have a policy on research data management. In the majority of cases, the policy is of very recent date.

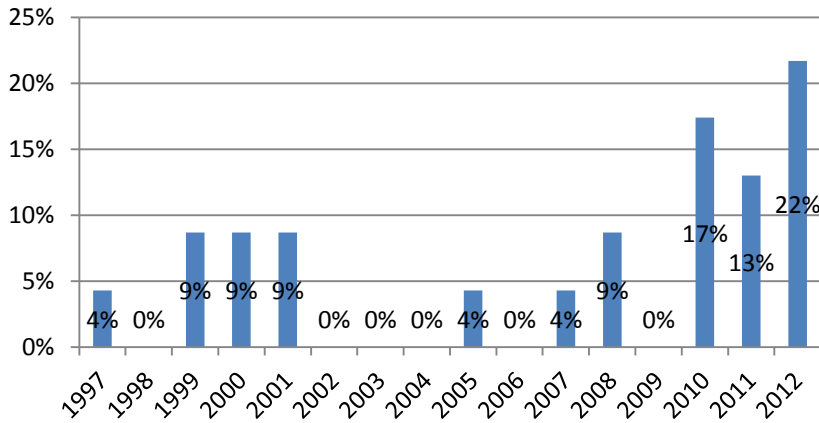


Figure 9: Years when policy was established (n=23)

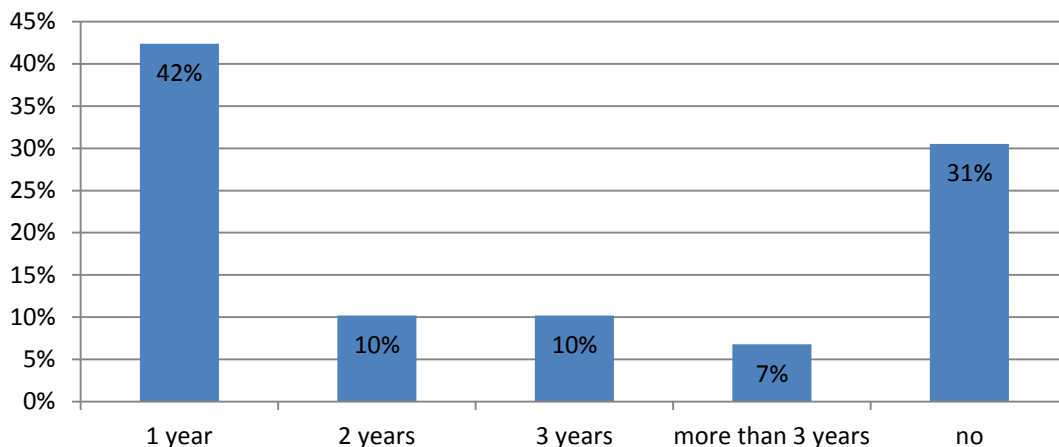


Figure 10: Years when policy was established (n=23)

Of those institutions without a policy (n=59), 30% have no intention of creating one, and 42% intends to create one in the next year.

Drivers for establishing a Data Management Plan

The drivers for establishing a Data Management Plan for the institutions with and without a policy on research data management are shown in Figure 11.

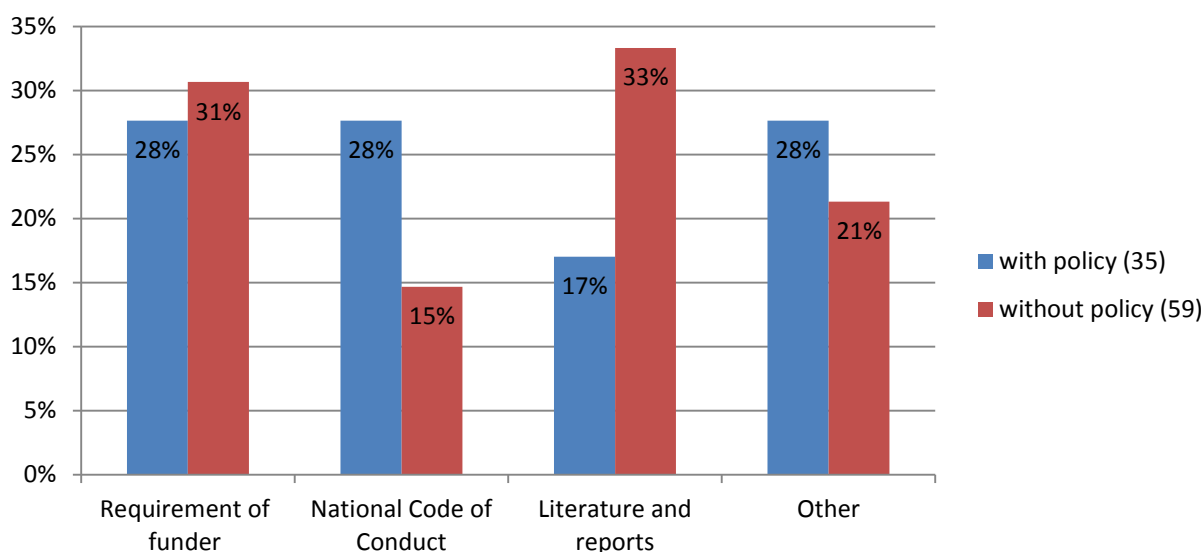


Figure 11: Drivers for establishing a Data Management Plan (n=94)

Other drivers for establishing a Data Management Plan are legal requirements, such as a national law on archiving or a university's code of conduct. Support for future research was mentioned several times. Some institutions see the collection, management and preservation of research data as their responsibility or primary task. The effectiveness and progress of the research community will be enhanced by having access to as wide a range of shared knowledge as possible.

Guidelines such as the Research Ethics Data Storage and Retention Guidelines⁵⁴, the United Nations Convention on Biological Diversity⁵⁵, and the Global Biodiversity Information Facility

⁵⁴ Canterbury Christ Church University, research Ethics and Governance Advisory Note available at Research Ethics Data Storage and Retention Guidelines

⁵⁵ Available from <http://www.cbd.int/> <http://www.cbd.int/>

Memorandum of Understanding (which aims to make biodiversity data openly and universally available)⁵⁶ were pointed out as reasons to set up a Data Management Plan.

For organisations without a policy, the drivers are not very different. Frequently mentioned drivers were legal requirements, project deliverables for eligibility for funding, and funding bodies' policy frameworks. It is notable that academic demand, a general commitment to Open Access and the importance directors and data management staff attach to data management can obviously drive the establishment of a Data Management Plan. These reasons were given several times. Other reasons were the challenge of sharing resources in order to gain efficiencies and more effective collaboration. The need to provide effective support to researchers by ensuring that an agreed set of data management requirements and expectations are in place between the researcher and host institution can also be listed in the aforementioned category of drivers. Other stated drivers for a Data Management Plan were IT security regulations, security concerns, and critical remarks stemming from the Audit of the State Accounts.

Elements of policies for data management

Figure 12 shows elements of the policies on research data management:

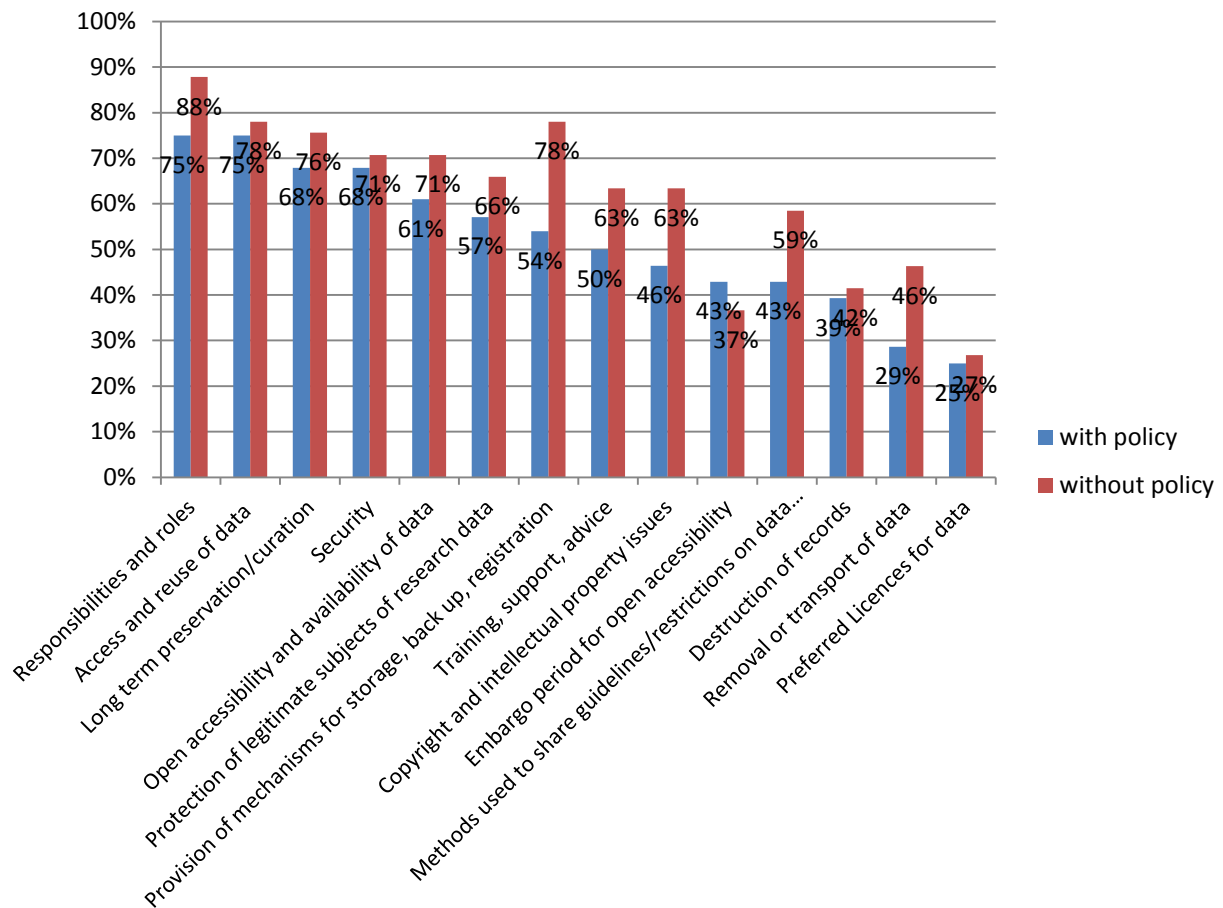


Figure 12: Elements of policies for data management (n=94)

⁵⁶ Available from <http://data.gbif.org/terms.htm?forwardUrl=http%3A%2F%2Fdata.gbif.org%2Fdatasets%2F>

The table below (table 4) shows the most important elements of such a policy according to the organisations:

Responsibilities and roles for management of research data	42
Training and support	29
Access and re-use	25
Security	24
Long term preservation/curation	23
Open accessibility and availability of data	17
Copyright	12
Protection of legitimate subjects of research data	9
Share data	5
Re-usability	4
Support service	4

Table 4: Most important elements of policies for data management

The results of the survey show that 15% of the organisations prescribe a Data Management Plan. The following figure shows the section headings of these plans:

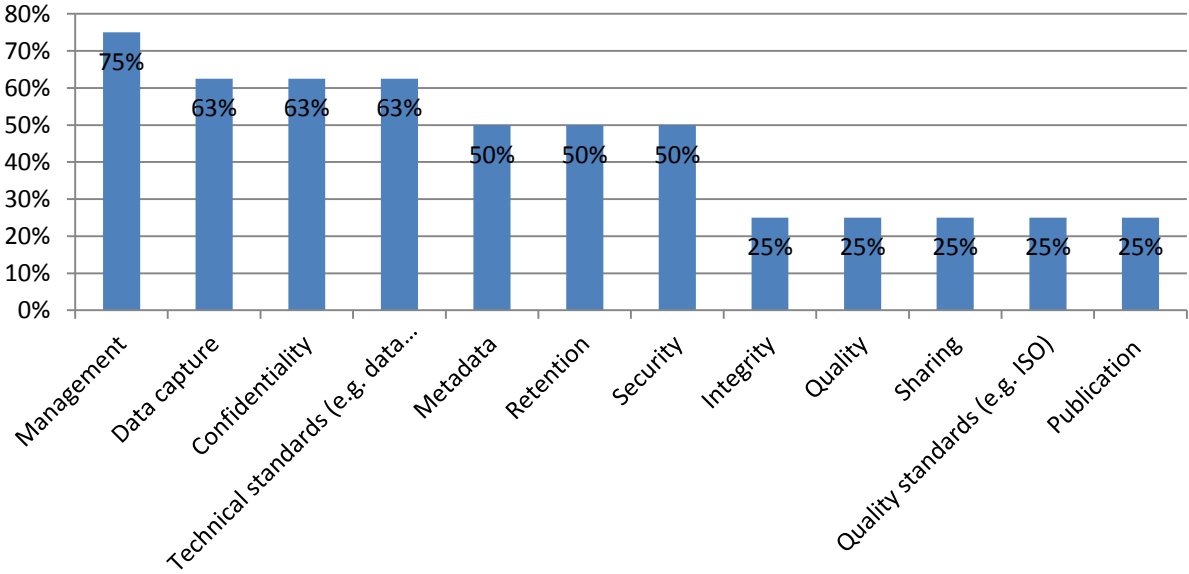


Figure 13: Headings of Data Management Plans (n=10)

3.3.2 Recognition

Incentives to deposit data

The survey formulated the question concerning incentives to deposit data as an open question. The findings are therefore more difficult to enumerate, but they provide a good overview of the incentives in place at some of the institutions surveyed.

Of the organisations that responded, 25% created incentives to deposit data. Several have set up a basic infrastructure or a common database and offer archives and recording or centralised services for researchers. One university mentioned providing free data publishing services to academics and translations of metadata into English so as to improve the visibility of their research data. This category of incentives also covers assistance with long-term archiving and a secure trusted

repository. Universities encourage staff and faculty to use SharePoint for storing documents that need to be accessed by many people across the university. Other incentives have been created to help researchers document data and keep track of data users and publications based on archived data.

There are also financial incentives. One of the organisations provides partial funding for creating necessary metadata and for applying standard formats.

Some of the incentives are indirect in that they offer Digital Objects Identifiers (DOIs) and supporting research infrastructures (technology services).

One data archive participates in projects with data producers and funds "small data projects" to ensure that the data is deposited correctly. The archive offers training and outreach activities but also provides access to thousands of scientific datasets, e-publications and other research information.

Re-use of data

Thirty-nine per cent (39%) of the organisation facilitate the re-use of data in-house. This was again formulated as an open question, so the answers vary considerably.

Organisations facilitate data re-use mainly by creating data repositories. There are varying levels, ranging from centralised (national) storage (data warehouse or a central 'Bank of Knowledge') to shared disk space or an infrastructure (intranet) to make data traceable and to make it easy to send a request for data. One institution has several operational databases, a dynamic copy of which is regularly updated to a data warehouse. From there, data – all types of data from different systems – can be queried for new purposes.

The organisations also mentioned websites for accessing and re-using the data. Other solutions that make re-use of data feasible are lists of available datasets in-house or on the intranet. Sometimes metadata is published publicly on the web to encourage discovery.

Another re-use strategy is to link the data to other sources. For instance, institutions store research data in the institutional repository to link them to articles or to combine different datasets (proprietary data from different datasets combined with the data of other institutions or research).

In the Netherlands, one way of re-using data is through the online archiving system EASY.

The data repositories are supported by formal or informal general or open access policies.

Of those organisations that facilitate re-use of data in-house, only one has outsourced this task. Of those that do not facilitate the re-use of data, 13% outsourced this to a national data archive or data centre. At European level, life scientists depend on services provided by European Bioinformatics Institute (EBI)⁵⁷ (for data deposition and distribution).

Licence for making research data accessible

Nearly a quarter of the respondents (22%) require a certain type of licence to make research data accessible. Some institutions have created their own licence whereas others recommend a standardised licence, for example one of the Creative Commons licences, Open Data Commons licence or the Global Biodiversity Information facility (GBIF)⁵⁸ Data Sharing Agreement. CLARIN⁵⁹ has produced model licences for unrestricted academic use and for restricted use (where more detailed application procedures are required).

⁵⁷ Available from <http://www.ebi.ac.uk/>

⁵⁸ Global Biodiversity Information facility Data Sharing Agreement available from <http://data.gbif.org/tutorial/datasharingagreement>

⁵⁹ Common Language Resources and Technology Infrastructure available from <http://www.clarin.eu/external/index.php>

Public health datasets carry their own restrictive licences due to both intellectual property and confidentiality issues. The precise nature of the licences varies.

One respondent indicated that there are numerous licences depending on the nature of the data and funders. This respondent is working on a project that rationalises these into three boilerplate licences to cover most eventualities.

Another respondent mentioned needing the approval of the data protection ombudsman for research.

Several respondents indicated that the permission of the data owner was required to use most research data. One institution said that it requires a licence to collect data, as most authors are not willing to share their information.

3.3.3 Institutional research data support

Responsibility

Figure 14 shows who is responsible for data management within the institution.

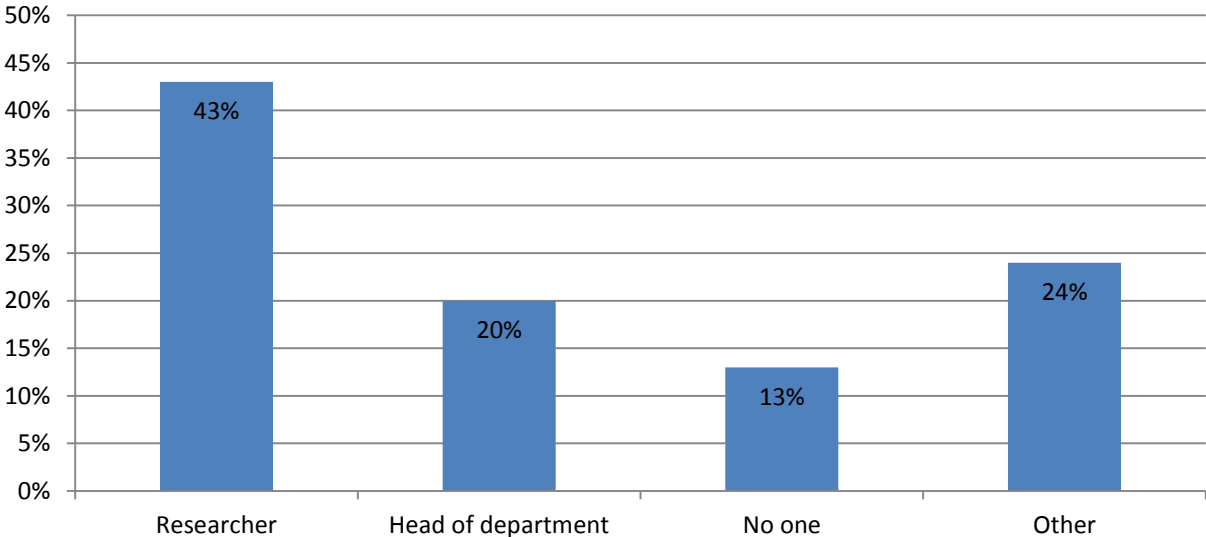


Figure 14: Responsibility for managing data (n=93)

In most institutions, a wide range of different staff members or departments are responsible for data management alongside the researcher/principal investigator or head of department. Universities have installed a cross-service, data management group or a university research committee. Examples of specially appointed staff are directors, heads of IT, librarians, information managers, data managers or specialists.

Support for research data management

Over half of the institutions (58%) support research data management. Support is provided in various ways. One organisation summed up all the support it gives its researchers: the university develops training material (so far mostly for humanities researchers); articles on methods and software tools for research data management which can be added to the university's Research Skills Toolkit; a research data repository; a 'database-as-a-service' system; a 'data finder' tool; a university-customised tool to assist with data management plans; and advice and customised web solutions for data management provided by computing services.

Not all institutions provide such extensive support. Some assist academic staff by adopting policies on records management, protection and/or retention. Providing information, guidelines and training are other mechanisms that organisations use to support their researchers. Information is

occasionally aimed at junior scientists. One organisation trains its researchers by stimulating the use of a data management plan.

Infrastructure support is provided at differing levels. This may simply mean providing disks, or it could also involve building a comprehensive support service and infrastructure, with advice, guidance and templates. One university responded that it has just purchased a research management system which will enable researchers to store publication data and professional activity. The system is also linked to the institutional repository, which is an open access database of published and unpublished works by faculty and researchers. The same institution also supports researchers in using SharePoint for data management and sharing. One organisation offers a persistent storage solution.

The establishment of a central databank or data dissemination system is another support mechanism. One university has set up an online research data management hub.

Assistance is also provided by data management groups, which take care of archiving data and making them accessible. In one case, the researcher can consult the IT department.

Support is often embedded within the governance and research office of the research organisations, but national data archives and centres are established to provide further backing. Organisations from Ireland, Finland, the Netherlands and the United Kingdom refer to support from these national bodies.

Experts to assist in data management

Many of the institutions (40%) employ an expert who can help researchers manage their data. These employees have different areas of expertise. Computer engineers, IT security officers, data protection officers, data and/or technical archivists, and data storage experts assist researchers. Respondents occasionally mentioned that the library offered support; IT staff were mentioned more frequently. One university stated that assistance with managing data is part of the library's specialist work. Libraries also have metadata specialists on staff.

The expertise of the specialists lies in data documentation, data management, data preservation and digital preservation. Expert knowledge is also available on protocols and the technical details involved in collecting samples, on relevant data portals and metadata standards, on national and international collaboration, and on developments in data management. Know-how related to statistics and programming, applications and hardware is also mentioned. Legal expertise with respect to intellectual property and data management licences is available as well.

Training in data management

Over one third (34%) of the organisations provide training for researchers in data management. Universities deploy a variety of tools to do this. These can be divided into online training via data management websites and online material, and face-to-face training such as lectures, day or week courses (not only on data management but also on statistical programs and project management in the research context). Several universities offer student research fellows short summer courses. One university mentioned that the course is non-mandatory for PhD students. Other means of communication include visits, printed publications, and information services via e-mails or telephone.

Staff training is not always planned. One university gives mostly ad-hoc advice: instructions are issued to the principal investigator prior to the research project. Another institution organises training on request for research staff and students.

Who actually provides training may depend on its purpose, and it may be provided by multiple individuals. At local level, training is provided by an information system service team who advise on software, system administration and hardware aspects of data management. The library service offers guidance on curation and publication of data. Information services provide information security training.

One organisation responded that it is in the process of planning the curriculum. Most likely it will start with 'how to use existing standards and vocabularies when producing or working with the data to enhance the interoperability of the separate data sources'.

Regarding the FIN-CLARIN⁶⁰ repository, seminars and courses are given in which users learn how to prepare and deposit their materials into the FIN-CLARIN repository.

Awareness about the re-use of data

Organisations were asked whether they raise awareness about the possible re-use of the data. Nearly half of them (48%) do. Their answers reveal a variety of standard public relations instruments: personal communication, websites, meetings and seminars, written flyers and briefing papers. Respondents often indicate that awareness is raised at an informal and personal level. Additionally, social media and websites are used, but personal interaction with researchers is an important way of raising awareness among this target group.

Various institutions organise meetings with departmental data representatives or start discussions with research groups. Within the research groups, the professors often encourage such discussion. One institution mentioned that they have people speak at some (but not all) introductory sessions for new researchers.

Training sessions or workshops or annual seminars are other awareness-raising mechanisms. Other means of making researchers aware is to publicise guidance materials, flyers or briefings. Presentations and symposia were mentioned several times as a means of disseminating insights into research data management.

One university applies a research data management hub that explains the benefits of re-use and provides assistance tools.

Those institutions that do not (yet) raise awareness about data re-use, were asked what kind of incentives should be created to do so. Policy recognition and compliance with policy were incentives that more than one institution mentioned. One university saw the requirement of open access when funding research as a solution.

Incentives were thought to depend on the stakeholder, but among the examples given were the increased impact of research; new discoveries; time and cost savings owing to proper management of institutional assets; time savings owing to the re-use of existing research datasets. All these incentives would ultimately lead to increased funding for research.

Other incentives cited were formal credit in citations and the inclusion of data publications in the university's list of publications/annual academic achievements report. Also mentioned was the inclusion of funding for data management in research grants and consideration given to data management in project audits and reports. One respondent replied that a reasonable percentage of the budget should be allocated for the purpose of managing research data and improvement of the same.

A central management facility and training would also be an incentive. Another driver is the possibility that researchers can store their data securely.

⁶⁰ The FIN-CLARIN consortium is part of the international CLARIN project. It aims to ensure that all the researchers in Finland are able to easily find and access all the European CLARIN-compatible language resources available from <https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/KielipankkiFrontpage>

3.3.4 Institutional tools and analysis

Infrastructure

Nearly half of the organisations (49%) have an infrastructure for storage, management and access for research data management. Respondents mentioned that they use a variety of file storage systems (NAS, SAN), library systems and services, such as a wiki. The institutional repository is also used for finished datasets (published in publications). Some of those are being extended to support the curation of datasets. There is often a common IT infrastructure. One organisation stated that all staff have a H drive allowance of 2GB standard, which can be increased on request. It also has a distinct data repository (triple backed up) at multiple locations, which is currently used for 'big' imaging data. A university in Ireland has an easy-to-use, web-based application that allows its researchers to maintain, update and publicise their research profiles. The system also provides a framework for meeting the requirements of funding bodies with respect to disseminating research output. The university's data warehouse provides convenient access and reporting capabilities relating to student research data.

Some institutions have a mixture of centralised networked storage and local resources in large research units. Others only provide storage infrastructure centrally, but not all of the units necessarily use this. Some departments have their own data management framework developed by their own IT groups. There is no solution for the institution as a whole. It was mentioned that a lightweight tool for local data management (structuring, metadata description) is to be offered to all groups, but that is still in the development stage.

The data centre for technical universities in the Netherlands has two copies of the datasets, which are stored outside the centre at different places. The data centre accepts all kinds of formats. Some formats are converted to NetCDF for better search possibilities. On request, the centre will build data labs, a research environment to share/access data during the study or project. They can be built with limited access.

Some countries have well-established national data archives, such as the Social Sciences and Humanities data archive, DANS in the Netherlands, or the CSC IT Centre for Science in Finland.

Of those organisations with an infrastructure for research data storage, management and access, the majority (89%) are aware of the needs of researchers with regard to research data management. One respondent commented that some academic disciplines are understood better than others. Nearly two thirds (65%) of the organisations with an infrastructure have adapted it to the specific needs and wishes of the researchers.

Universities choose diverse routes to investigate and examine the needs of researchers. Among these routes are active and on-going consultation and dialogue with the research community (and their representative bodies within the institution). Another frequently mentioned route was consultation with (and instruction from) the relevant national and European funding bodies.

Sometimes awareness is based on the experience of in-house researchers. Knowledge about researcher needs and requirements is gained by means of interviews with the liaison librarians and by researchers contacting the institutional repository. Other information about researcher needs is acquired in face-to-face contact and by collecting feedback from talks and seminars and courses. Some universities have account managers who are in frequent – and in some cases, daily – contact with researchers.

Several organisations conducted surveys, sometimes followed by interviews and collaboration with selected pilot partners, to investigate their current data management situation. The surveys addressed researchers or institute leaders across the entire higher education sector.

Payment

Figure 15 shows who pays for data storage and data management (n=30).

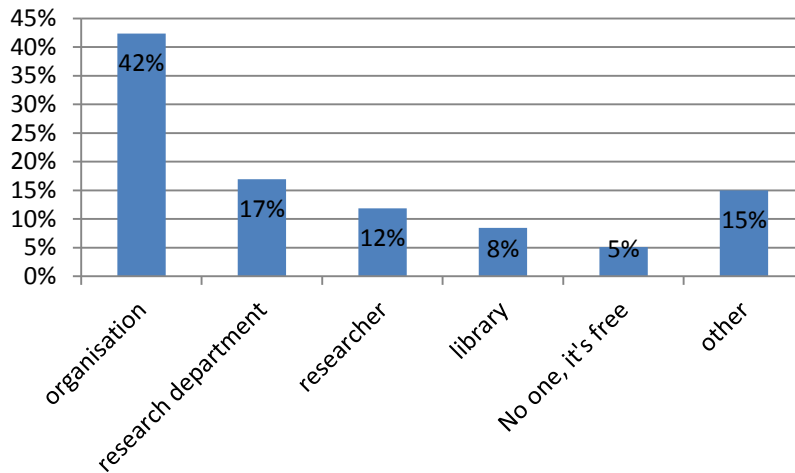


Figure 15: Funding of data storage and data management (n=18)

Other sponsors of data storage and data management beyond the options listed are the government via the Ministry of Education and Culture, the university's IT department, or the research funder when primary research data management is part of the grant.

One respondent stated that payment depended on the size: small datasets are free; for a big dataset (>50 GB), the research department and the library pay for storage and management.

With over a quarter (28%) of the organisations without a data infrastructure, some responses included comments concerning the cost of not storing the research data.

Tools and utilities

Tools and utilities for storing and managing research data are supplied by nearly half of the organisations (49%). There are various examples, nearly always on an institutional level. In the Netherlands, however, there is a remote access system for handling the data at the national level, just as there is in Finland, where FIN-CLARIN supplies tools for building metadata and annotating materials and where CSC (which is part of FIN-CLARIN) supplies other tools.

Universities provide disk space or access to a server, offer archiving possibilities, have their own databases and build on campus-licensed database management system (Microsoft SQL and Oracle tools). Sometimes the universities' own traditional storage facilities or data warehouse are used for storing and managing the data.

Tools that universities use are the GBIF Integrated Publishing Toolkit, Nesstar publisher, International Household Survey Network (ISHN) National Data Archive application (NADA) catalogue, Open Metadata Framework, Questasy data dissemination system, Dropbox, current research information system tools and SharePoint.

Many organisations are designing, developing and testing local management tools for structuring and metadata description. One university mentioned that most of the tools are still under development, although some data is now stored in earnest in their data repository. That university is testing its online Database as a Service system, which should be formally launched in the summer of 2012. The university is also working on a Data Finder tool that should be ready to launch in early 2013.

Central research data management system and facilitation for research data collection

Over a quarter of the organisations (28%) have a central research data management system. Slightly fewer (21%) facilitate the collection/capture of research data and their subsequent

publication. DANS in the Netherlands developed EASY, which can, in theory, store all sorts of data and all formats: text, databases, spread sheets, images and sound.

An Irish university hosts an institutional publications repository in addition to the IRIS system. This is a freely accessible online database holding its researchers' digital research outputs. This repository supports the preservation, dissemination and promotion of the university's research and showcases the wide range of research taking place within the university.

One respondent envisages a workflow front-end using a web-based GUI, e.g. SharePoint or similar. The workflows will model the typical data administration tasks that a researcher has to do.

Tools for finding and organising research data

Of the respondents, 17% of the organisations have a tool that finds and organises research data. The questionnaire did not permit respondents to list what tools were used.

Conclusions

The number of institutions with a research data management policy is growing. The majority of these policies were defined in the past two years. Of those institutions that do not have a policy, 42% intend to define one in the coming year. Of the organisations without a policy, 30% have no intention of defining one.

The most important drivers for creating a Data Management Plan are the requirements of funders, a national code of conduct and the influence of literature and reports. Other drivers include guidelines from diverse national and international organisations, legal requirements and, security issues.

The drivers for institutions with or without policies do not differ very much. For institutions that do not have a policy, academic demand or a commitment to open access in general can drive the establishment of a Data Management Plan.

The data management policies cover a range of elements, of which the most important are to define the responsibilities and roles relating to research data management, training and support, access and re-use, security, long-term preservation/curation, and open accessibility and availability of data. For institutions without a policy, the most important elements do not differ all that much except that the provision of mechanisms for storage and back-up is considered as important as access and re-use of data.

Of the organisations that responded, 15% prescribe a Data Management Plan. The section headings in these Data Management Plans vary, but management, data capture, confidentiality, technical standards, metadata, retention and security are found in 50% or more of the Data Management Plans.

A quarter of the organisations create incentives to deposit data. Several institutions have created an infrastructure for researchers or provide services to their academics regarding data publishing or long-term archiving. Some institutions also create financial incentives.

Institutions facilitate the re-use of data mainly by creating data repositories. Other solutions are to offer information about available datasets or linking services to assist in tracing the data. Data repositories often are supported by open access policies.

Universities generally require a certain type of licence to make research data available. Some institutions have created their own licence, but if they use a standard licence they recommend either one of the Creative Commons licences, the Open Data Commons licence, or a standard licence recommend by a discipline.

In most cases the researcher is responsible for managing the data. If it is not the researcher who is responsible, it is usually the head of department. 13% of the respondents indicated that no one bears responsibility.

Over half the institutions support research data management. They do so in many different ways. Most common are to adopt policies and provide infrastructure support. Assistance by data management groups and training are also common mechanisms.

Researchers receive help in managing their data from specialists with all kinds of expertise regarding different elements of data management. It is usually IT staff who provide support. Not only do they offer assistance but they also provide training by deploying a variety of online and face-to-face tools and instruments. Nearly half the institutions raise awareness of the possible re-use of the data. They use all sorts of public relations instruments, both formal and informal. Social media are also used.

Those institutions that do not raise awareness about data re-use see compliance with and recognition in policy as incentives to do so, as well as the requirement of open access when funding research. Incentives for data re-use depend on the stakeholder, but include increased impact for

research leading to more funding, formal credit for citations and inclusion of data publications in the university's list of publications, and research funding for data management.

Nearly half the organisations already have an infrastructure for research data storage, management and access. Various file storage and library systems are used, often with a common IT infrastructure. Institutional repositories are also deployed to store finished datasets. The Netherlands, the United Kingdom and Finland have well-established data archives for storage. Of those universities with an infrastructure, many have adapted that infrastructure to the specific needs of researchers. They investigate those needs in active and on-going consultation with the research community. The tools and utilities are once again mixed, but they are nearly always supplied on an institutional level.

More than a quarter of the organisations have a central research data management system. Of the organisations, 17% have a tool that finds and organises research data.

4 Publishers

4.1 Research methods and response

Publishers with a known contact person received an invitation by e-mail to participate in the survey. Contact persons from publishers' associations distributed the questionnaire among publishers.

After several weeks only six publishers had responded. As all other publishers already received the invitation and reminder, it was decided to contact them by telephone. When a first attempt was not successful, at least two more attempts were made to reach the contact person or somebody else to answer the questions. In the cases the right person in the organisation was contacted, an offer was made to fill out the survey together.

It proved to be difficult to get hold of the right persons in the organisations that could answer the questions of the survey, if there was somebody knowledgeable on this subject at all. Many respondents stated they did not know enough on this subject to be able to answer all the questions. Others requested the link to the online survey. However, this did not lead to much response: they often stopped filling in the online survey after only a few questions. Eventually only ten (10) publishers participated in the survey, mostly from the UK, but also from The Netherlands, Finland and Lithuania (see table 5).

COUNTRY	RESPONDENTS
UK	4
LITHUANIA	2
USA	2
FINLAND	1
THE NETHERLANDS	1

Table 5: Publishers

4.2 Summary of findings

Most publishers have heard about data research policies, but have not developed a policy yet or are still at the beginning of the progress of developing it. At the moment, research data policy is for most publishers still an abstract subject and it will take more time before publishers can say something meaningful on this subject.

Asked whether they have a data policy, five publishers responded that they do not. The data policies of the remaining two publishers cover only certain journals. Most Life and Earth Sciences titles have data policies of varying strictness and complexity. One publisher was aware that other subject areas have at least had discussions about data policies, but was unsure what point these discussions had reached. It was explained that the journals that operate in data-intensive disciplines do have data policies. These policies concern issues like as depositing data in public repositories, submitting accession numbers with manuscripts, the type of data allowed in supplements to the article, and how to represent data within the article. Authors need to be able to provide links to the data underlying their primary research articles, preferably for readers, and most certainly for reviewers. Difficulties in providing these links could result in articles not being accepted for publication. Sometimes direct discussions with editors lead to a solution (in cases involving commercially sensitive information or other licensing restrictions).

Asked whether they require research data to be submitted together with the article, two publishers said yes and five said no. Two publishers also required the entire dataset to be archived.

Two of the publishers consider peer review of datasets as part of the article review procedure. In one case data is published separately from the article and is available on demand. In the other case the data are published supplementary to the article. Publishers do not ask the researchers to update their data.

Where the data is published, the DOI is used as persistent identifier for citing datasets.

The location of the citation with regard to the article varies. Publishers said that they tend to place the citation in the reference list if the data is not an integral part of the article, but other variations are possible. They indicated that accession numbers are usually mentioned inline in the article and that links can appear in footnotes. One publisher responded that they only cite data if the author chooses to do so in either the reference list or in the acknowledgements. This publisher answered that it has the facility to cross-link data DOIs with PANGAEA⁶¹, but that it has not yet populated these links.

None of the respondents employs a standard for data. Only one has created an incentive for depositing data, namely encouraging all authors to do this and making it part of its best practice guidelines.

Authors encounter problems when depositing research data because there are no trustworthy repositories in many disciplines or if they do exist, they lack a DOI registry system to link to and from publications. Of those authors who do deposit data, 40% have had problems sharing their data for legal reasons and for reasons of potential misuse. The respondent referred to the PARSE.insight report⁶².

Not depositing data can lead to non-publication. One publisher mentioned this possibility. None of the publishers indicated that they have special provisions in their publishing contracts with authors concerning the publication of data.

Asked what feasible intervention would be best for improving the research data management skills and knowledge of researchers, the publishers gave various different answers. One said that a powerful intervention would be to require all underlying data to be available in a trustworthy repository when the manuscript is accepted for publication. Another powerful intervention might be the publication of datasets and data papers via peer-reviewed journals.

4.3 Conclusions

Not enough publishers have filled in the questionnaire to draw valid conclusions. We have therefore only summarised the publishers' responses here.

The responses show that data policies are not established instruments yet. Existing policies require authors to provide readers and reviewers with links to the data underlying their primary research articles. Authors need to submit the entire datasets when submitting the article. Not depositing their data could mean non-publication. Authors do not need to update the dataset. Publishers generally do not create incentives for depositing data.

Publishers use a DOI as a persistent identifier for citing the data. There are different ways to cite the data. Publishers do not utilise a standard for data.

Authors might have different reasons for not depositing data: legal reasons and fear of misuse, but also the lack of trustworthy and DOI-compliant repositories.

Two powerful interventions were suggested to improve authors' research data management skills and knowledge: requiring all underlying data to be available in a trustworthy repository at the time an article is accepted, and publication of datasets and data via peer-reviewed journals.

⁶¹ PANGAEA: Data Publisher for Earth & Environmental Science. The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing geo-referenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions, available from <http://www.pangaea.de/about/>

⁶² PARSE.insight. Insight report

5 Validation of interventions

Chapters 3 and 4 show the interventions that stakeholders have taken to stimulate and improve the management of research data. Especially regarding the policies of research institutions it is unclear whether those approaches chosen actually improve the skills and knowledge of researchers to use and benefit from the current and future scientific e-infrastructure for their research.

In order to verify the effectiveness of the different interventions of the research institutions, several researchers in the partner countries of the SIM4RDM project have been interviewed. This was done on the basis of a pre-prepared questionnaire similar for all the persons to be interviewed. In this questionnaire the main elements that were found in research data management policies were listed and the researchers were asked to give their opinion about those components. Other questions that were raised regarded their drivers for writing a data management plan and the type of interventions an institution should take to support them in their data management activities.

The results have been scored and corrected for meaningful response, for example if the respondent indicated there was no policy within the organisation or the respondent didn't know, the answer to whether the respondent was familiar with the contents of the policy were disregarded. For some questions the answers of some respondents were deemed unusable for quantitative analysis. For multiple choice questions, each answer is scored 1 point (questions 2.1, 2.2, 2.3, 3.1, 3.2, 3.3 and 5.1). For questions to prioritise interventions, each answer is scored 1 to 5 points: the item indicated as most important is scored 5 points, second most important 4 points etc. (questions 2.4).

For categorised multiple choice questions, the scores per category are calculated by summing the individual terms (scoring 1 point each) which are then weighed for a theoretical maximum score ($N_{\text{terms}} \times N_{\text{respondents}}$) for the given category (question 4.1). For those questions to prioritise categorised interventions, each answer is scored 1 to 3 or 1 to 5 points and for a theoretical maximum score for the given category ($N_{\text{terms}} \times N_{\text{respondents}} \times \text{pointS}_{\text{max}}$) (questions 4.2, 4.3). Given the number of questions and the number of respondents, the results are subject to over-dimensioning. This means that the results of this survey should be used indicative only.

A total of 12 researchers were interviewed, mostly from the UK but also from The Netherlands, Ireland, Sweden, and Finland, coming from different disciplines and covering a part of the broad spectrum of positions in a scientific environment. The questionnaire was submitted to a head of a department, four professors, one senior lecturer, a former researcher now director of a library, a young research engineer, and researchers. They work in the department of sociology, physics, scientific computing, bio sciences, biochemistry, and learning technology, psychology and law. Two of the interviewees are a director of a library.

To the question whether the organization has a research data management policy for managing research data 5 respondents answer affirmative, 6 negative and 1 interviewee said he doesn't know. All the researchers employed at an institution having a research data management policy know the elements the policy consists of.

In the opinion of the interviewees all elements that were mentioned in the questionnaire, should be part of a policy for managing research data but some elements were stated more repeatedly. The five elements that were mentioned most frequently in order of importance are responsibilities and roles for managing data, mechanisms for storage, backup, registration, deposit and retention of research data, access and re-use of data, open accessibility and availability of data, and long term preservation and curation. Destruction of records and preferred licences for data are seen as the least important elements of a research data management policy.

Three (3) institutions have a mandate for writing research data management plans. In the opinion of the respondents the main drivers for writing a data management plan are the requirements of a funder or publisher. Furthermore, respondents were driven by many other reasons to write a

research data management plan. Several times it was brought up that it is the researcher's code of conduct: it is part of good researchers' practice and it makes research easier and more efficient. Researchers are bound to engage in research in an ethical and sound manner. Good data management is an important element of this. Another motive that was given was the fact that research is financed with public money and therefore the outcomes must be used as best as possible. In line with this motive is the motivation that it is a requirement to retain data as it is generated on behalf of others as a service. One respondent sees self-protection as a driver for writing a plan; in the case of leaving employment the data remain with the institute. One respondent thinks it unnecessary to write a plan if this only is demanded by a funder without any benefit in the future for the researcher itself.

Four (4) of the respondents have never written a research data management plan.

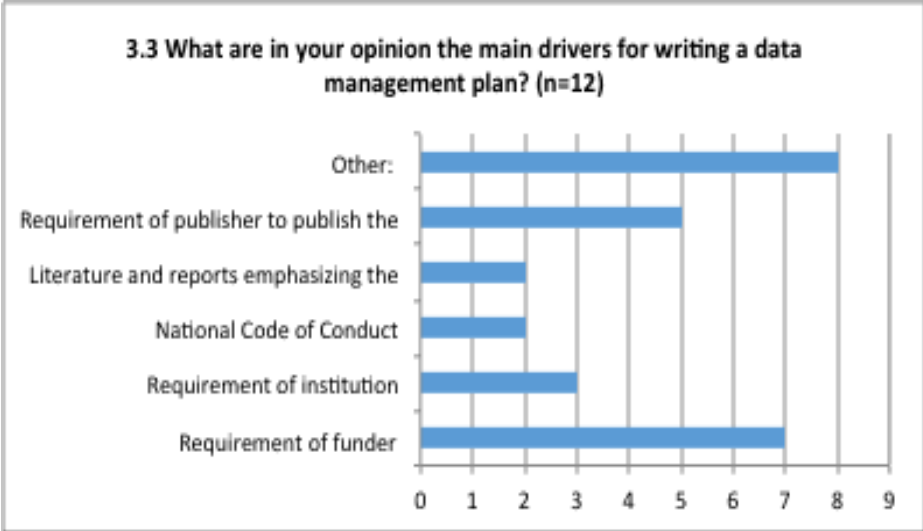


Figure 16: Main drivers for writing a data plan (n=12)

To the question what type of interventions would support a researcher in its data management activities, interventions regarding data infrastructure were raised most often, followed by recognition, institutional data support, support and training, and incentives for depositing data. Storage facilities and tools to identify and organise research data scored the least.

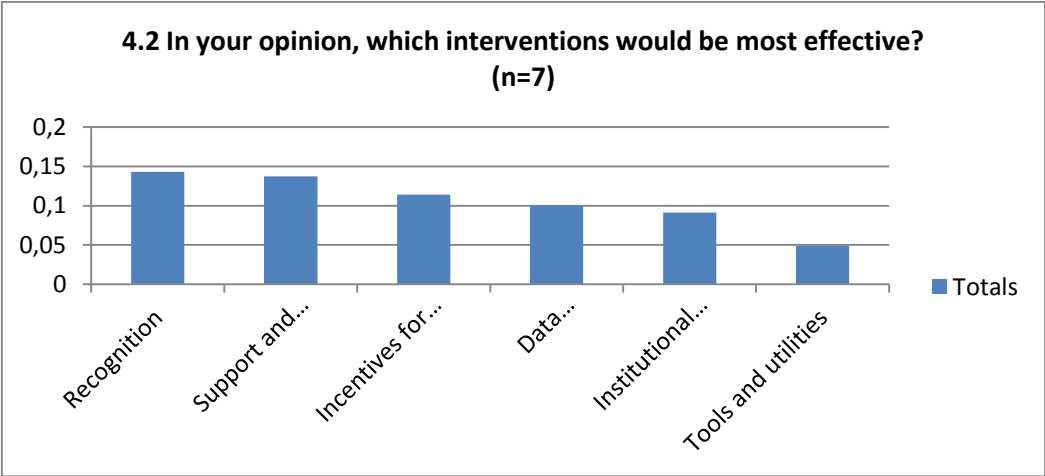


Figure 17: Interventions that would support a researcher in its data management (n=7)

In the opinion of the respondents recognition, support and training, incentives for depositing data, and institutional data support are the four most effective interventions. Recognition should be more than just thanking a researcher for sharing his/her data in the end of a paper. Regarding support and training, experts to assist with research data management are seen as most forceful. Other support could be career development training and ad hoc face-to-face support. Concerning the incentives for depositing data, linking data with other resources for analysis, data repositories and several others were listed. One respondent stated explicitly that there should be no financial bonus for the researcher to deposit data.

For depositing data, proper storage with metadata is critical. At the same time data repositories must do something for the researcher and must be easy to use. It seems that most current systems work for the benefit of the institution and not for the user. A question was raised about depositing data; how are other researchers made aware of existing data?

On the subject of institutional data support, comprehensive support services are liked most. Other support services could be adequate training, help from a national body for smaller institutions, and in-house IT services. It was mentioned that the only system that would work is targeted support at the disciplinary level. The overheads in financial models should include for disciplinary-level data management support.

Researchers consider tools and utilities the least effective intervention. Especially flyers, meetings, seminars and presentations are seen as not very effective. In the category tools and utilities, recognition and institutional data support don't score very well either.

6 Workshop findings

To augment the survey results, a workshop organized by which took place in Zandvoort, the Netherlands on the 25th and 26th of April 2012. Participants were identified during the survey and by the Steering Committee for the SIM4RDM project. The goal of this two-day workshop was to:

1. Produce insight for possible interventions and how to establish those;
2. Analyse interventions already in place, opportunities and similarities;
3. Reflect on which interventions might be successful to reach the set goal;
4. Establish which interventions should be part of the framework.

Proceeding from general issues outlined below, participants of the workshop talked about data management in all its aspects. They discussed the subject of speeding up the progress of science by making available freely all the research data so that researchers would gain countless new insights without first spending years on fieldwork. Combining and comparing data sets would strengthen the reliability of findings. And at a more prosaic level, publishing data together with the article may allow other fellow researchers to better assess the findings, even to the point of discovering fraud. In some disciplines, such as genomics, it's already standard practice that complete data sets are stored in institutional and other repositories, open to other researchers or even the wider public.

The big question the participants asked themselves is how to get researchers to cooperate. This led to a numbers of problems:

- It's extra work to upload their data in a format that other people can understand and use. This may even require new or specialized skills.
- Researchers may not want competitors to profit from all the hard work they've done collecting the data.
- Some or all of the data may be sensitive to privacy issues.
- There is no real recognition (yet) for collecting data in itself.
- It's a leap of faith for researchers to grasp the possibilities of open data for their own work.

On both days of the workshop, discussions took place in small groups after an introduction of the perspective of a representative of one of the stakeholders.

One important insight that rose from the workshop is the need to cast the SIM4RDM net wider. There are far more actors involved than just researchers, funders, research institutions and publishers. Examples include various national and international organisations and bodies, and the editorial boards of scientific journals. Data centres and infrastructure providers (e.g. libraries) are relevant but play differing roles. Scholarly societies (such as Britain's Royal Society) were also often mentioned as representing researchers, although this point of view was not universally supported.

Young academics are a prime target group, although the participants could not settle on a common definition of this group. Generally, the SIM4RDM project needs to settle on common definitions of the various roles, for example 'data producer', 'national body' and 'end user'.

Differing interventions were identified for the different stakeholders. For researchers, these were the use of a shareable licence and recognition for citing data. For research societies, they included developing common practices for citing data from articles or teaching young researchers about data quality. Infrastructure providers could intervene with common data formats for preservation and storage, tools and utilities. There may be opportunities for data scientists to maintain good quality data catalogues. Funders could use a wide range of interventions. The participants came up with policies to use shareable licences for research data, promoting usage of standards and practices, requiring the use of approved data repositories in cooperation with editorial boards and defining policies of sustainability.

The main theme was that policies from funders, institutions and editorial boards may influence researchers to use the principle of 'share and share alike'. Data management and data citation should be part of regular scientific practice, but awareness and understanding will first need to be raised among researchers.

7 Conclusions

The aim of SIM4RDM (Support Models for Research Data Management) is to enable researchers to effectively utilise the emerging data infrastructures by ensuring that they have the knowledge, skills and support infrastructures necessary to adopt good research data management methodologies. Funders of research and research institutions can support and assist researchers in managing their data by requiring certain interventions or by installing tools and employing expert staff. To explore what interventions funding agencies or research institutions can use to ensure that researchers are supported in the management of their data, the project team conducted an online survey among those organisations. The outcomes of the online survey provide an overview of the interventions currently used by the stakeholders. Based on these outcomes, the team has produced recommendations for interventions.

7.1 Funding organisations

Nearly half of the research funding organisations have a policy covering data management. A quarter of these require a Data Management Plan as part of the grant application, in most cases for all disciplines. For funders, the most important drivers for a Data Management Plan are better research and science and support for open data and access. According to funders, a Data Management Plan should address data acquisition, use, re-use, storage and protection, and the rights of ownership. Just over one third of the organisations designate a specific organisation for preservation. In general no term for preservation has been identified.

7.2 National bodies

Several national bodies have been established to coordinate research data management activities, mostly funded by the government. National bodies are more often than not project organisations. Their most important tasks are to coordinate access to tools and materials, to establish long-term preservation and sharing practices, to assist in the policies of research funders, and to create infrastructures.

7.3 Research institutions

The number of organisations with a research data management policy is growing. Not all institutions without a policy intend to create one, but among those that do, 42% intend to do so in the next year. The drivers to create a Data Management Plan are not very different for institutions with and without a policy. They create one because of the requirements of funders or due to a national code of conduct. They are also driven by literature and reports. Other reasons for establishing plans are guidelines issued by various national and international organisations, legal requirements, or security reasons. For institutions that do not have a policy, academic demand or a general commitment to open access can drive the establishment of a Data Management Plan. Researchers are driven by many other reasons to write a research data management; more than once they brought up that good data management is part of the researcher's code of conduct and that researchers are bound to engage in research in an ethical and sound manner. Another motive is the fact that research is financed with public money and therefore the outcomes must be used as best as possible.

Data Management Plans generally include the following elements: definition of responsibilities and roles involved in research data management, training and support, access and re-use, security, long-term preservation/curation and open accessibility and availability of data. Institutions without a policy consider storage and back-up mechanisms more important than institutions with a policy.

The five elements that are considered most important by the researchers questioned are responsibilities and roles for managing data, mechanisms for storage, backup, registration, deposit and retention of research data, access and re-use of data, open accessibility and availability of data, and long term preservation and curation. Destruction of records and preferred licences for data are seen as the least important elements of a research data management policy.

Only 15% of the respondents prescribe a Data Management Plan. The section headings of Data Management Plans vary, but management, data capture, confidentiality, technical standards, metadata, retention and security are found in 50% or more of the Data Management Plans.

To encourage researchers to deposit data, the organisations have set up infrastructures for them. They also provide services to their academics for the publication of data or long-term archiving. Re-use of data is facilitated mainly by building data repositories or linking services to find the data. Simply providing information about available datasets is another incentive. Data repositories are often supported by open access policies.

Universities generally require a certain type of licence to make research data available. Some have created their own licence, but a standard licence is also recommended.

In most cases the researcher is responsible for managing the data; if that is not the case, the head of department is often responsible.

Universities support research data management in many different ways. Most common are the adoption of policies and infrastructure support. Other common instruments are assistance by data management groups and training. Researchers can rely on specialists with all kinds of expertise to assist them. Most of the support is provided by IT staff. Institutions also provide training using a variety of online and face-to-face tools and instruments. Nearly half of the institutions raise awareness about the possible re-use of the data; they employ all sorts of public relations instruments at formal and informal levels, as well as social media. Incentives for re-use are instruments that lead to more funding, formal credit for citations and inclusion of data publications in the university's list of publications. Another incentive is funding for research data management. Researchers consider tools and utilities the least effective intervention. Especially flyers, meetings, seminars and presentations are seen as not very effective. On the other hand, comprehensive support, adequate training, experts to assist with research data management are seen as forceful. Furthermore, career development training and ad hoc face-to-face support are seen as effective interventions.

Nearly half of the organisations have an infrastructure for storage, management and access for research data management. A variety of file storage and library systems are used for this. A common IT infrastructure is often used; the institutional repositories are thus part of the infrastructure. The infrastructure is adapted to the specific needs of researchers. To investigate their needs, the institutions engage in active and on-going consultation with the research community. The tools and utilities are mixed, but they are nearly always supplied at an institution-wide level. More than a quarter of the organisations have a central research data management system and 17% have a tool that finds and organises research data. Data infrastructure storage facilities, tools to organise data and central data management systems are often regarded as effective.

7.4 Publishers

Not enough publishers have completed the questionnaire to draw valid conclusions. Their responses have merely been summarised here. It appears that data policies have yet to become established instruments among publishers. The policies that do exist require authors to provide readers and reviewers with links to the data underlying their primary research articles. Authors need to submit the entire datasets when submitting the article. Not depositing them could mean non-publication. Authors do not need to update the dataset. Publishers do not utilise a standard for data but they use a DOI as a persistent identifier for citing the data. Data are cited in different ways.

The lack of trustworthy and DOI-compliant repositories may prevent authors from depositing data; other reasons may be related to legal issues and fear of misuse.

Two powerful interventions were suggested to improve authors' research data management skills and knowledge: require them to make all underlying data available in a trustworthy repository when their article is accepted, and publication of datasets and data via peer-reviewed journals.

8 Recommendations

8.1 Funding organisations

Funders turn out to be the main driver for institutions to establish Data Management Plans. It is recommended that funding organisations

- Encourage researchers to create a data management plan at the stage of the definition of the project proposal;
- Allocate part of the research grant for data management purposes as funding may be an important incentive for data re-use;
- Help to advance research data management by designating data centres for physically storing research data;
- Issue instructions for writing the research plan and define clear deliverables for research projects;
- Consider and identify what elements of a Data Management Plan will actually assist researchers in their data management.

8.2 National bodies

A Code of Conduct is another driver for creating Data Management Plans. It is recommended that national bodies

- Take the lead in drafting such a code. A Code of Conduct should outline responsibilities or proper practices. It may guide the decisions or procedures on the part of individual researchers or research institutes in the area of data management. A Code of Conduct could also initiate or structure debates about adequate practices and models in a broader European context;
- Suggest and supply appropriate tools and best practices. Appropriate tools that already exists in some European countries could be adapted to prevent other national bodies from re-inventing the wheel;
- Play an active role in standardising regulations and procedures and defining conventions for data citation;
- Are established with designated tasks regarding co-ordinating activities related to research data management. In cases this is not possible another option might be liaising with a pan-European research data management body;

8.3 Research institutions

It is recommended that research institutions

- Develop data management policies that contain elements supporting scholars and scientists in their management. Institutions should ensure that the following elements are addressed: the responsibilities and roles involved in research data management, training and support, access and re-use, security, and long-term preservation/curation;
- Develop and provide a template or distribute examples of successful data management plans, in order to support researchers to produce a data management plan;

- Build a trustworthy research data infrastructure and create workflows for data publishing and archiving;
- Help researchers to decide on research data management tools and processes that best serve their research purposes by giving them information about tools and good practices available both at the national and at the European level;
- Help bridge the gap between data and publications by encouraging researchers to link their data with their publications and give them credit for this;
- Provide their staff with guidance on intellectual property issues and licensing of data, in order to raise awareness amongst their researchers regarding about important research data related aspects.

8.4 Publishers

Because of the limited response from the publishers valid conclusions about their dealings with data are hard to draw. Desk research showed that authors can submit underlying research data with their publication but that publishers stated authors are responsible for those data. Publishers stated that they don't have arrangements in place to preserve the data. From the minimal responses it is clear that publishers are still in the beginning of formulating data policies.

It is recommended

- To establish a dialogue with publishers or publishers' associations about establishing data policies. Possible elements of those policies could be creating incentives for depositing data, standardisation of persistent identifiers for citation of data, and requirements for trustworthy repositories to deposit the data.

8.5 Recommendations from the workshop

The two-day workshop made clear some of the definitions used in defining the roles within the environment of research data management are not unambiguous. To better clarify which stakeholders are meant it is suggested

- To settle on common definitions of the various roles as for example 'data producer', 'young academics' or 'end-user'.

The outcome of the workshop also showed that that the setting for research data management is broader than first thought. There are more actors involved than initially identified. Thus it is recommended

- To bring in other stakeholders in as well. Possible examples are international organisations and bodies, editorial boards of scientific journals, data centres and infrastructure providers.

It is further recommended that

- Interventions are tailor-made so as to make them more effective in enhancing researchers' skills and knowledge regarding data management.

Different stakeholders need different interventions, hence the recommendation that stakeholders should look carefully which interventions they deploy. In particular it is recommended that

- Funders use a wide range of interventions;

- Researchers use a shareable licence;
- Research societies intervene with the development of common practices for citing data from articles or teaching young academics about data quality;
- Infrastructure providers intervene with common data formats for preservation and storage, tools and utilities.

Finally, an important recommendation is

- To use policies from funders, institutions and editorials boards to influence researchers to apply the principle of share and share alike. Data management and data citation should be part of regular scientific practice but awareness and understanding will first need to be raised.