



Project no. 269977

APARSEN

Alliance for Permanent Access to the Records of Science Network

Instrument: Network of Excellence

Thematic Priority: ICT 6-4.1 – Digital Libraries and Digital Preservation

D35.1 EXEMPLAR GOOD GOVERNANCE STRUCTURES AND DATA POLICIES

Document identifier:	APARSEN-REP-D35_1-01-1_0
Due Date: Submission Date:	2013-12-31 2014-01-17
Work package:	WP35
Partners:	BL, CERN, CINI, CSC
WP Lead Partner:	CSC
Document status URN	Released urn:nbn:de:101-20140516264

<u>Abstract</u>: This report summarises the level of preparedness for interoperable governance and data policies based on both desktop research on selected data policies and online survey conducted during this study. It is important to understand what current data policies address and if they miss out on important topics, such as specific requirements for data preservation. This will give an indication on the possible impact of such data policies on the individual communities and allows recommendations to be drawn up to guide forthcoming policies. This report concludes with selected recommendations that should be taken into account when drawing up data policies concerning digital preservation.



Delivery Type	Report					
Author(s)	Suenje Dallmeier-Tiessen (CERN), Mariella Guercio (CINI), Heikki Helir (CSC), Patricia Herterich (CERN), Kirnn Kaur (BL), Artemis Lavasa (CERN) Juha Lehtonen (CSC), Riina Salmivalli (CSC)					
Approval	David Giaretta / Simon Lambert (coordinators)					
Summary	This report summarises the level of preparedness for interoperable governance and data policies based on both desktop research on selected data policies and online survey conducted during this study. It is important to understand what current data policies address and if they miss out on important topics, such as specific requirements for data preservation. This shall give an indication on the possible impact of such data policies on the individual communities and allows recommendations to be drawn up to guide forthcoming policies. This report concludes with selected recommendations that should be taken into account when drawing up data policies concerning digital preservation.					
Keyword List	Data Management, Data Policy, Digital Preservation, Governance					
Availability	Public Public					

Issue	Date	Comment	Author
0.1	2013-12-02	First version to be submitted for internal review	BL, CERN, CINI, CSC
0.2	2013-12-12	Second version to be submitted for external review	CSC
0.3	2013-12-18	Third version based on the review by René van Horik (DANS), Brian Matthews (STFC), Rob Baxter (EPCC), and Michelle Lindlar (TIB)	CSC
1.0	2013-12-20	Final version	BL, CERN, CINI, CSC

Document Status Sheet



Project information

Project acronym:	APARSEN
Project full title:	Alliance for Permanent Access to the Records of Science Network
Proposal/Contract no.:	269977

Project coordinator: Simon Lambert/David Giaretta				
Address:	STFC, Rutherford Appleton Laboratory Chilton, Didcot, Oxon OX11 0QX, UK			
Phone:	+44 1235 446235			
Fax:	+44 1235 446362			
Mobile:	+44 (0) 7770326304			
E-mail:	simon.lambert@stfc.ac.uk / david.giaretta@stfc.ac.uk			



CONTENT

1	INTRODUCTION	6
	1 MOTIVATION AND BACKGROUND	6
	2 OVERVIEW OF GOVERNANCE STRUCTURES AND DATA POLICIES	6
	3 Methodology	7
2	COVEDNANCE STRUCTURES AND DATA DOLICIES	0
4	GOVERNANCE STRUCTURES AND DATA POLICIES	0
	1 GOVERNANCE AND FUNDING	8
	2.1.1 Memory Institutions	9
	2.1.1.1 Governance structures	9
	2.1.1.2 Funding	10
	2.1.2 Universities and research institutions	10
	2.1.2.2 Funding	10
	2.1.3 Research funding agencies and policy makers	11
	2.1.3.1 Governance structure	11
	2.1.3.2 Funding	12
	2 DATA POLICY AND DATA POLICY ADHERENCE	13
	2.2.1 Memory institutions	13
	2.2.1.1 Data policy adherence	13
	2.2.1.2 Data poincy adherence	14
	2.2.2 Onversities and research institutions	14
	2.2.2.2 Data policy adherence	15
	2.2.3 Research funding agencies and policy makers	15
	2.2.3.1 Data policy	15
	2.2.3.2 Data policy adherence	16
	5 DATA REQUIREMENTS, PRESERVATION AND RE-USE	1/
	2.3.1 Memory institutions	17
	2.3.1.2 Open access and re-use	18
	2.3.1.3 Preservation requirements and time frames	18
	2.3.1.4 Metadata requirements	19
	2.3.2 Universities and research institutions	19
	2.3.2.1 Preserved data	19
	2.3.2.2 Open access and reuse	19
	2.3.2.4 Metadata requirements	20
	2.3.3 Research funding agencies and policy makers	20
	2.3.3.1 Preserved data	20
	2.3.3.2 Open access and reuse	20
	2.3.3.3 Preservation requirements and time scale	20
	4 OWNERSHIP AND RESPONSIBILITIES	21
	2.4.1 The memory institutions	21
	2.4.2 Universities and research institutions	22
	2.4.3 Research funding agencies and policy makers	22
2		
3	ΙΠΕ Ουκνεί	23
	1 SURVEY RESULTS	24
	2 Results in Selected Countries	28
4	RECOMMENDATIONS AND CONCLUSIONS	
Ŧ		
	1 GOVERNANCE AND FUNDING	30
	2 DATA POLICY AND DATA POLICY ADHERENCE	31
	DATA REQUIREMENTS, PRESERVATION AND RE-USE	
	4 OWNERSHIP AND RESPONSIBILITIES	
5	REFERENCES	33

APA Contraction of the second second

6	ANNI	EXES	35
	6.1 I	Legal Requirements	35
	6.1.1	Memory institutions	
	6.1.2	Universities and research institutions	35
	6.1.3	Research funding agencies and policy makers	35
	6.2 (QUESTIONNAIRE	



1 INTRODUCTION

1.1 MOTIVATION AND BACKGROUND

Research data has been a central product of the research lifecycle for centuries. Today, it is very much the centre of attention for researchers and other stakeholder groups. Society and policy demand open access to research data funded by public money. Reproducibility of research, secondary analysis and enabling the linking of research results (publication) with related data sets are the main drivers for preserving and sharing data. Today, research data is considered a valuable scholarly object for which the producers should gain credits and acknowledgement.

The APARSEN project focuses on data preservation, and thereby on the main topics of "trust", "sustainability", "usability" and "accessibility". All of them are key requirements for the future reuse of datasets. Governance and data policies contribute to how research data is handled in research communities, organisations and institutions: they are part of the current lively "carrot and stick" discussion. Policies are issued by different kinds of organisations, from funding bodies to community representatives to national initiatives.

There is a need for an analysis of the current state of data policies in order to make recommendations for new organisations on how to draw up data policies. It is important to understand what data policies address and if they miss out on important topics, such as specific requirements for data preservation. Understanding this gives us an indication of the possible impact of such data policies on individual communities and allows recommendations to be drawn up to guide forthcoming policies.

This report studies some current data governance structures and policies in Europe and beyond. Different topics – for example metadata requirements or licensing – were identified by the project and surveyed through desktop research and an open survey.

The crucial concepts at the basis of this report are commonly used in the information environments but are not sufficiently investigated with reference to the differentiation of meanings behind them. Governance and policy, in particular, are key terms which are used in many areas for identifying peculiar functions. Governance is commonly interpreted as set of principles, procedures and policies for the management and government of institutions but also complex phenomena. More specifically, the term is referred to the capacity of guaranteeing a coordination action within a complex system where multiple and particular positions and interests are involved without losing the efficiency and the quality of the final decision. The term policy can be defined with the same degree of flexibility as a principle or rule to guide decisions and achieve rational outcomes, a statement of intent. In this report both terms are referred to the specific area of data management, access and preservation.

1.2 OVERVIEW OF GOVERNANCE STRUCTURES AND DATA POLICIES

This report summarises the level of preparedness for interoperable governance and data policies among surveyed communities and organisations. The original plan was to give targeted guidelines at national level on how to improve governance and data policy interoperability. However, based on research, it was evident that giving different recommendations at national level was not necessary.

Rules, procedures, general statements are present in institutions to support access to, and preservation of, their digital heritage. Because of the complexity of the digital environment, governance structures as well as data policies are needed. The crucial questions include governance structures, methodology for the implementation of policies, the analysis of specific (formally recognized) policies for data/records management and preservation. They also cover their specificity with reference to national and European strategic plans and to international standards in specific domains: memory institutions (libraries, archives, and museums), universities and research institutions, and research funding agencies and policy makers.

These policy documents focus on issues relevant to preservation management, even if not explicitly related to the preservation process itself, such as: legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, accountability, sustainability, retention of policy management mechanisms and repository restrictions, and metadata requirements.



This gives rise to the first fundamental challenge of defining the scope and focus of our research: what is a policy and, more specifically, what is a policy for digital preservation?

Policies are developed practically in the real world to support common behaviour and to ensure normalized practices. In ICT environments, where heterogeneity and dynamicity challenge the capacity of organizations to cooperate, the practical implementation of technical rules cannot be left without a general strategy and a standardized approach. Consistent vocabularies and clear definitions help make these efforts sustainable and efficient specifically if the aim is to provide the basis for comparison and for developing common recommendations. It is not by chance that a taxonomy approach has been at the centre of all recent projects in this field, even if common taxonomies have not yet been identified, other than in the form of a simple framework and list of activities for building a policy: for example, in the investigation of Madeline Sheldon [35] (see Table 1).

Access and use	Preservation Model/strategy
Accessioning and ingest	Preservation planning
Audit	Rights and restriction management
Bibliography	Roles and responsibilities
Collaboration	Security management
Content scope	Selection/appraisal
Glossary/terminology	Staff training/education
Mandates	Storage, duplication and backup
Metadata and documentation	Sustainability planning
Policy/strategy review	

 Table 1: Digital Preservation Policies Taxonomy (Sheldon)

A complex taxonomy for policies related to information management has recently been published by the International Foundation for Information Technology (IF4IT), with the aim of defining types and categories of policy [16]. This defines a "linear" list of 991 types – in many cases partitions of a policy – identified because they are "important to many IT organizations, IT Professionals, and the Business stakeholders they provide services for".

All these exercises make evident the need for a more comprehensive approach to, or at least a more consistent analysis of the increasing number of policies required for building, managing and accessing a digital archive.

Because existing analyses are not conclusive and the technical and theoretical framework is not sufficiently developed to provide enough elements to support a methodological approach on governance models and data policies, the APARSEN project has undertaken, as a first step in this direction, a survey to better understand the actual mechanisms currently in place for supporting interoperability and facilitating data access.

1.3 METHODOLOGY

Our approach was twofold: firstly we conducted desktop research with the aim of collecting and investigating current policies and rules as well as general recommendations proposed by policy makers or funding agencies; secondly, we conducted an online survey through which we gathered input from various stakeholders. The desktop research was organized in specific sections: governance, policy, data preservation and re-use, ownerships and responsibility, data requirements, funding, data policy adherence, and other information. The structure of this report was defined by analysing some examples of data management policies available at national and international level and includes all the main critical aspects that emerged in the course of the analysis with the aim of providing a sort of comprehensive state of the art. The third chapter of this report is dedicated to the web survey results. From both these analyses a list of recommendations has been collected, organized by type of institutions and their main related functions.

As the number of published data policies is currently limited, and digital preservation is tightly connected to data policies, the list of data policies reviewed was complemented by a list of more



general policies (listed in Table 2). These were analysed with particular focus on their applicability to research data management.

Table 2. Policies created by the following institutions were analysed

Memory institutions
The National Archives UK
National Archives of Estonia
The National Library of Wales
The National Archives of Canada
National Archives of Australia
National Museum of Australia
Archives New Zealand and National Library of New Zealand
National Library of Australia
National Library and National Archives of Norway
Parliamentary Archives UK
Universities and research institutions
Data Information Specialists Committee – UK (DISC-UK)
German Research Foundation
Ministry of Education and Science of Spain
Swedish Research Council
Sapienza Digital Library
Research funding agencies and policy makers
Biotechnology and Biological Sciences Research Council (BBSRC) (UK)
European Commission
Canadian Heritage Information Network
Ministry of Science & Technology of India
Agenzia per l'Italia digitale
JISC (UK)
Ministry of Education and Culture of Finland

2 GOVERNANCE STRUCTURES AND DATA POLICIES

To help with our review of policies we created a categorization scheme to demonstrate the common features of specific organization types. This categorization enables us to highlight the most relevant aspects and tendencies among the preservation policies, and where elements remain unplanned within the organizations and where there exists a need for further development work and support.

To ensure quality and consistency when comparing information on policies developed by different types of organizations, we categorized the organizations in question as memory institutions, universities and research institutions, and research funding agencies and policy makers. Because the categorization used here has been developed merely to help clarify the policies analysed in this context, it may not be applicable for other purposes. It also provides a contextualized assessment of the relations between the rules considered here and the relevant international or national legal frameworks that can reveal interesting characteristics in different environments.

The number of analysed cases is limited, and so the analysis cannot provide significant statistics or draw more than general conclusions; however, some tendencies are clear and reveal a reasonable picture of the current status of digital preservation policies in use currently.

2.1 GOVERNANCE AND FUNDING

A governance structure provides an organisation or institution with a framework describing clear roles and responsibilities, linked to the organisation's or institutions' strategic objectives. Organizational policies and procedures drive the definitions of roles and responsibilities within the governance structure such that it can fulfil the decision making and leadership duties assigned to it. Links to policy



then become transparent and manageable. The aims of our analysis were: to review whether the governance structures in place would allow for data policies to be implemented; and to consider whether a methodology for implementing such structures was identified or stated within the policy documentation. This analysis proved to be a difficult exercise as the documents reviewed (which related to data polices) did not always describe explicitly the governance structure which would allow the policy to be implemented.

Sufficient funding is a crucial element for any organization to exist as an entity. Depending on the type of organization, the funding models may be quite diverse. Where policies are linked to other organizational strategies it may be assumed that the work undertaken within the policy is also funded as part of within the core business of the organization. For the organizations reviewed, funding relevant to policy is tied intrinsically to the organizations overall funding and thus to the links between policy and organizational strategy. The analysis focused on whether the activities which support data policies were carried out as core business activities, so that additional funding was not required, and whether this funding was dedicated to the actions and activities described within the policy.

The analysis is provided below for governance structures and funding by sector: memory institutions, universities and research institutions, and research funding agencies and policy makers.

2.1.1 Memory Institutions

2.1.1.1 Governance structures

The National Archives (TNA) [24] is the UK government's official archive. Their role is to collect and secure government records, preserve them for future generations and make them accessible and available. Their expertise in effective management of information makes them a valuable resource for a wide range of users, from the 250 government and public sector bodies to members of the public. As well as adhering to an internal preservation policy they also provide guidance and advice on a number of topics including digital preservation policies. The guidance on policy for digital preservation archives discusses the need for, and need for support for, digital preservation and is intended to improve the governance of digital preservation through the development of digital preservation policy.

Within its policy documentation, the National Library of Australia (NLA) [28] defines some principles to be carried out by the Library. These principles cover aspects such as cost, staff, access, collection development, risks, audit, reports, format, migration, and more. The principles provide some insight into the governance structure although they do not explicitly define one.

The Digital Preservation Strategy of the Archives and the National Library of New Zealand [3] provides guidelines for the digital preservation of heritage content in New Zealand, and information regarding the need for collaboration between the Archives and the National Library is stated. The strategy document provides some information regarding the commitment of both organizations for carrying out this strategy, and also states that responsibilities have to be settled. In this instance the governance structure is not defined explicitly within the document.

The National Archives of Canada [19] provides a document with instructions and policy for preservation activities. The document states that all the policies will be carried out within the Archives, according to the roles and responsibilities, and will be monitored as described in the policy. Clearly this provides a governance structure within which the organisation can implement its policy for preservation.

The Ministry of Culture of Norway [31] aims to have a joint preservation and dissemination system between the National Library, National Archives and organisations within the museum sector. The internal preservation policy of the National Archives of Estonia [27] considers it important to continue with and develop cooperation in the field of preservation with other memory institutions and organizations on the international, national and institutional level. It is not clear whether governance structures are provided for within the documents reviewed.



2.1.1.2 Funding

At The National Archives (TNA) in the UK [24], activities supporting the preservation policy are part of its core business and apply to all areas of the organisation. Links to other policies, processes and best practice are stated within the policy document. The National Library of Australia's documentation [28] states that the organisation will "work with governments to develop legislative and funding frameworks to enable cost-effective preservation".

The National Archives of Canada only mentions that funding opportunities can be explored in relation to acquisition [19].

The Ministry of Culture of Norway [31] in relation to the National Library and National Archives states that in principle, expenses should be covered within the institutions' own budgets. The Ministry states that a common national solution will require extra equipment purchases to increase the repository capacity at both the National Archives and the National Library. There will also be a need to develop models for organising and funding dissemination services.

The strategy presented in the joint vision of both the Archives and National Library of New Zealand [3], requires that the organizations develop policy documents to create a "practical framework within which the long-term preservation of digital content can take place". There does not seem to be another information regarding funding within these policy documents.

2.1.2 Universities and research institutions

2.1.2.1 Governance structure

The Priority Initiative 'Digital Information' [2], a joint initiative of the members of the Alliance of Science Organisations in Germany, states that each organisation should be responsible for the implementation of each of the individual policies. The governance structure provides for working groups which are coordinated by the steering committee, in which all partner organisations are represented. Monitoring the progress of the workings groups is considered the central task of a steering committee. Therefore, each working group reports to the steering committee which develops proposals for the further advancement of existing core activities, for the initiation of a new or for the possible closure of an existing core activity within the initiative. The members of the steering committee are listed in the document.

The Ministry of Education and Science of Spain, and the Spanish Geological and Mining Institute reference the international Joint Committee on Antarctic Data Management (JCADM) where the governance structure is provided [21]. The Centro Nacional de Datos Polares (CNDP, the Spanish National Centre for Polar Data) is responsible for some tasks (defining types and formats of data) and it will give support to researchers by providing the means for them to manage their data.

The report "Policy-making for Research Data in Repositories: A Guide (DISC-UK DataShare)" [8] provides a guiding document based on the recommendations of three different organizations (the Universities of Edinburgh, Oxford and Southampton). The guide can be used by digital repositories as a decision making and planning tool. It established institutional data repositories and related services at the universities stated. The guide can also contribute indirectly to articulate the benefits of data management practices, specifically it assists in the development of policies about data management and preservation.

2.1.2.2 Funding

In terms of funding, the Alliance of German Science Organizations states [2] that, "Cost structures play a crucial role in the installation of suitable research data centres and research infrastructures, both national and local. The setup, development and operating costs for selected research data centres in different disciplines will be evaluated in order to obtain a clearer picture of the cost factors for local and distributed data centres in the field of data management". The Alliance partner organisations are "particularly concerned with coordinating their policies and practical decisions in relation to digital research infrastructures, pooling their financial resources in this area, and, where necessary, making additional resources available". The Alliance of German Science Organizations also states that, "the

Date: 2013-12-31



concrete aims of the working group are to promote funding for the 'golden road', define criteria for the adoption of open access publication fees, work to increase the content of open-access repositories and collaboratively support international open access infrastructures. Due consideration will be given to discussions and initiatives of Science Europe and the newly founded Global Research Council".

2.1.3 Research funding agencies and policy makers

2.1.3.1 Governance structure

The SIM4RDM (Support Infrastructure Models for Research Data Management), a two year ERA-NET project funded by the European Commission's Seventh Framework Programme (FP7), produced a report on the existing policies in various Member States and some international ones as well [23]. This report states that "existing pan-European and international grant programmes and institutions' policies" varied, meaning that the governance structures in place are different for every country, or every institution or organization. The report goes on to state that, "various national bodies have been established to coordinate research data management activities such as the 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. Most of these national bodies are funded by government, more often than not in a project structure". The report further recommends that, "national bodies could take the lead in drafting a national code of conduct which encourages the creation and use of data management plans; suggest, supply and adapt appropriate tooling to the national context; and, take an active role in data citation practices".

The document 'Putting Data into Use' reviewed from Finland [20] provides a roadmap for a future policy. The policy refers to large collaborations between ministries, research institutes and infrastructure providers and suggests the founding of a long-term coordination group under the Ministry of Education and Culture, which would supposedly provide a governance structure.

The European Commission's recommendation on 'Open Access to and Preservation of Scientific Data' [11] and the European Communication, 'Towards better access to scientific information: Boosting the benefits of public investment in research' [14] provide a national point of reference which was to be in place by the end of 2012, under which responsibilities would include (but not be limited to) coordinating the actions taken by the Member States in regard to this Recommendation and reporting back to the Commission for progress evaluation purposes. The report from the European Research Area Committee within the European Commission (EC), 'National Open Access and Preservation Policies in Europe' [12], states that each Member State should be responsible for implementing their own policies. Although, when asked about the role that respondents see for the European Commission in monitoring the progress by Member States'. Thus the EC's position provides a role in terms of some of the duties within an overall governance structure of a Member State.

The Beagrie study, funded by JISC [4], provides advice and guidance on getting started with setting up a digital preservation policy and includes the consideration of governance structures which are implemented by connecting digital preservation policy to the business drivers of the organization and analysing existing policies and strategies which are in place. With this approach, mappings can be made from high level strategic organizational policies to relevant areas of the digital preservation policy.

On a national level, the Canadian Heritage Information Network (CHIN), within the Government of Canada, provides a checklist for Creating a Preservation Policy [6]. The checklist is made up of a list of questions which help in creating a preservation policy for museums, although there is no further information on how to proceed. The methodology for implementing a governance structure is provided by the checklist which covers; Organization, Media, Formats, Metadata and Intellectual Property Rights, although there is no further information on how to proceed after completing the checklist of questions.



2.1.3.2 Funding

Date: 2013-12-31

The SIM4RDM [23] project recognizes the challenges arising from the lack of long-term funding at national level. The report on its survey states that nearly half of the organizations analysed have a funding policy covering data management. The funding of data storage and data management could derive from the organization itself, the research department, researchers or the library; or, in some cases, from the government via the Ministry of Education and Culture (or equivalent), the IT department (in the case of a university), or some other organization. Other responses to their survey mention that: some funds are to be reserved for support and storage of research data; a reasonable percentage of the budget should be allocated for managing research data (and improvements); and, partial funding for creating necessary metadata and for applying standard formats should exist.

The BBSRC Data Sharing Policy [5] states that it will facilitate and encourage data sharing by providing funds (through specific schemes) to support: the development of standards and software tools which enable data sharing; community resources and the facilitation of data sharing approaches in specific communities; and data sharing activities as part of the full economic cost of research projects. Support is also provided for relevant training activities. The following funding schemes are available to provide support for data sharing activities:

- Responsive mode research grants: BBSRC recognizes that data sharing has time and cost implications. Funding to support the management and sharing of research data (for example staffing, physical resources such as storage and networking capability) can be requested as part of the full economic cost of a research project.
- Special Opportunities:
 - Tools and Resources Development Fund: This is a pump-priming fund intended to support new and innovative projects, but also with provision to fund projects that promote data sharing or data standards
 - Bioinformatics and Biological Resources Fund: A fund designed to support the establishment, maintenance and enhancement of community resources required by bio-scientists
- Strategic Priority in Technology Development.

In Finland, a report by the Ministry of Education and Culture on, 'Putting Data into Use' [20], states that data should be available for free or with a small non-profit charge which should be taken into account by the research institutes. Such actions and activities require sustainable funding and it is suggested that the continuous, committed long-term development of financing models is undertaken, with the public research financiers playing a large role.

For the European Commission, the funding required should be a matter for each of the parties involved in the process (i.e., institution, organisation, or other). The only indications that the funding needed for implementing policies will have to become available are notes to the effect of: "making the necessary funding available for dissemination (including open access)", and "earmarking the necessary funds". The document produced by the European Research Area Committee within the European Commission, the 'National Open Access and Preservation Policies in Europe' [12], only makes references to the way specific countries deal with their funding issues. It states that "in some countries ..., some research organisations fund open access publishing ... to cover authors' fees" and that "in most European countries, direct government grants are the main financial mechanism for the funding of university libraries and public research centres". Additionally, financial support for journals and investments in dissemination are mentioned, as well as other agreements with publishers.

The Beagrie study [4], states that development and implementation of digital preservation policy are considered as core business activities which should be embedded into the organization's overall policies with clear mappings to other key policies. It notes that financial sustainability should be defined clearly as part of the overall organizational financial plan. For future work in this area, it recommends that (1) funders should utilize this study to evaluate and encourage best practice in terms of institutional preservation policies and procedures and their development in their funded programs and evaluations and (2) JISC should consider assisted take-up by funding case studies on how the



generic digital preservation policy in this study might apply in practice to specific activities, digital content or units within higher education institutes, for example institutional repositories.

2.2 DATA POLICY AND DATA POLICY ADHERENCE

Our analysis of data policies and adherence to them is presented below again by specific sectors: memory institutions, universities and research institutions, and research funding agencies and policy makers. In analysing adherence to the policies, our desktop research approach included questions regarding to the way adherence is monitored and by whom. This is a matter of great importance since following certain guidelines and protocols helps organisations focus on their aims and objectives.

The analysis of risk assessment or risk management procedures proved relatively challenging, since more than half of the policies did not integrate with any sort of risk assessment document. The same issue was also present in the study of whether policies were linked to any research or career assessment. Another critical question was that of the interoperability of the policies and whether they can be applied to similar institutions.

The authors of each of the policies and the purpose that those policies serve (preservation, access, reuse etc.), as well as the sector to which they can be applied are important components of the following analysis, because they provide context and valuable insight into the logic behind the policies and why their development was essential. The legal requirements for the policies mentioned in this chapter are listed in detail in Annex 6.1; it should be noted that there was no information available regarding legal requirements in nine policies (out of 25).

2.2.1 Memory institutions

2.2.1.1 Data policy

In this category, ten memory institutions were reviewed in relation to the data policies which they had adopted. These institutions are mostly libraries and archives, but also included are a museum and a ministry. Overall, eight institutions that have adopted data policies have designed them themselves. The policy of the National Library of Wales [29] was the result of work done by the digital preservation committee with the UK Parliament, the policy being written by an internal project team and a number of external reviewers. One of those external reviewers was Neil Beagrie, the director of consultancy at Charles Beagrie Limited. This company also wrote the policies with funding from JISC UK [4].

Of the ten policies, eight are meant to be applied to one or more departments of the organization which produced the policy [3, 19, 26, 27, 28, 29, 30, 32]. With reference to the two other policies [24, 31], in the case of Norway [31], the policy aims to be a national strategy in the archive, library and museum field. In the National Archives policy (UK) [24] it is mentioned that the archive also provides guidance notes and advice on preservation and/or management topics to other archives and to the public sector in the UK.

For all the memory institutions, preservation of materials in their possession (tangible or digital) is considered the main purpose of their policies. The policy of the National Library of Australia [28] specifies that their preservation tasks are focused on national heritage contents, information resources and digital collections. The Norwegian strategy [31] states that, besides preservation and dissemination, their efforts extend to digital protection and long-term storage of collections to make them digitally accessible. In addition, the National Archives (UK) [24] policy provides further guidance to highlight the need for organizations to adopt digital preservation policies, and for those policies to be linked to other policy areas such as governance structures.

Of the policies that were examined in this category, it is observed that six of them do not provide information concerning the issue of their application to similar organisations or institutions. However, three of the policies [24, 27, 31] can be applied to other similar organizations or institutions, either within the country of origin of the policy, or even at an international level, while one policy [26] simply states that interoperability is paramount.



2.2.1.2 Data policy adherence

Six cases do not contain any information on the person or people monitoring adherence to the policies [3, 26, 27, 28, 29, 32]. Four policies [19, 24, 30, 31] are overseen by departments or offices within the institutions that operate under the guidelines of each of those policies. In particular, the UK National Archives policy [24] mentions that preservation issues at all levels are monitored to ensure adherence to the policy by the collection care and digital preservation departments.

In eight policies, the responsibilities of data curation and preservation are exclusively an internal matter [3, 19, 24, 26, 27, 28, 30, 32]. In the Norwegian strategy [31], it is assumed that two collaborating national bodies (the National Library and the National Archival Services) will perform data curation and preservation, notwithstanding the fact that this was not confirmed. One policy in this category [29] had no instructions for the execution of curation and preservation activities.

In the Estonian policy [27], incentives are provided by supporting the development of the professional qualifications of the staff. The other nine memory institutions did not include any particulars in relation to any research/career assessment or incentives.

In five of the policies [19, 24, 27, 29, 30] it is discernible that the institutions or organisations conduct various kinds of risk assessment or risk management procedures, such as emergency or Disaster Contingency plans [19]. Five of the policies [3, 26, 28, 31, 32] did not incorporate such a document.

2.2.2 Universities and research institutions

2.2.2.1 Data policy

Of the five policies in this category, the German policy [2] was written by several contributing partners and the DISC-UK DataShare policy [8] was the result of the collaboration of three universities (Edinburgh, Oxford and Southampton). Two policies [34, 38] were designed by dedicated committees and one [21] by a ministry¹.

The Sapienza digital library policy [34] aims to be applied to all the departments, laboratories and research centres of the university. The other four policies [2, 8, 21, 38] are applicable at a national level to universities, institutions, research centres and others. The Spanish policy [21] in particular is a protocol with basic instructions for submitting, storing and making available research data. This Protocol can be applied to data, metadata and publications derived from the projects funded by the Spanish National Subprogram for Polar Research.

Furthermore, the above Spanish policy is aimed at data collection as well as storage and dissemination/facilitation of access, and covers building a digital repository with an emphasis on the interoperability of metadata. As far as the aims of the other policies are concerned, the German policy [2] aims to guarantee broad access to a wide variety of materials, to promote the international distribution and reception of German publications and research data, to ensure the long-term availability of materials and to support IT-based research. The DISC-UK DataShare policy [8] is intended to be used by institutions that are considering adding research data to their digital collections, and the purposes of the Sapienza digital library policy [34] are data sharing and open access. Finally, the case of the Swedish policy [36, 38] is noteworthy, since even though it is directed towards preservation, re-use and access; the Personal Data Act (see Annex 6.1) makes it illegal to use research data other than in specific projects, for which the data were collected.

In the DISC-UK DataShare policy [8] it is stated that where interoperability is concerned, the policy will be shared with the other academic institutions, specifically those involved any future implementation of similar services. The German policy [2] can be transferred to German universities, research institutions and/or other research organisations.

¹ The Spanish Ministry of Education and Science; specifically, by the Polar Archives within the National Centre for Polar Data, which is part of the Geologic and Mining Institute



2.2.2.2 Data policy adherence

In the German policy [2], adherence to the guidelines is overseen by a steering committee, in which all partner organisations are represented. Monitoring the progress of the working groups is considered the central task of the steering committee and therefore, each working group reports to the steering committee. Also, the Sapienza digital library policy [34] states that a strategic board is in charge of verifying the accuracy in the application of the guidelines and in the case of a lack of compliance the general procedures of the University are applied. The other three policies do not mention anything pertinent to this.

The Sapienza digital library policy [34] has allocated the data curation and preservation responsibilities to a strategic board and a scientific committee, while another, the Swedish policy [38] has assigned those responsibilities to the Swedish National Data Service. There were no remarks with regard to this question in the other policies.

The German policy [2] provides incentives by recognising and supporting the efforts for the provision of research data for further use, while the Swedish policy [38] gives incentives by providing funding for preservation tasks and so on, if a data management plan is provided. Of the policies in this category, three did not include any comments on any research or career assessment or incentives, and none of them incorporated a risk assessment document. In terms of future integration, in only one case (the Sapienza digital library policy [34]) is it reported that a risk assessment document would be integrated with the policy in the future.

2.2.3 Research funding agencies and policy makers

2.2.3.1 Data policy

It is common practice for research funders to institute certain requirements or guidelines concerning data policies. Of course, such requirements can differ from funder to funder; however, it is always expected that their funding recipients will comply with these requirements/guidelines in the data management policies that they will subsequently develop.

Of the three documents [4, 11, 23], two were already a meta-analysis of the policies currently in place across the EU Member States, as well as countries outside Europe, so it is difficult to identify one organization or group they apply to. On the other hand, the purpose of those policies is quite clear: the SIM4RDM study [23] was created to "produce an overview of possible interventions that have proved effective in supporting researchers in their 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 researchers". Accordingly, the 'National open access and preservation policies in Europe' [12] aims to improve dissemination of, and access to, scientific information, as well as the preservation of scientific information in Europe. The last policy relates exclusively to biologists, who receive funding from a certain organisation (BBSRC) [5] and seeks to encourage the researchers receiving this funding to share data.²

The BBSRC document does not omit to mention, however, that the policy can be transferable to other research institutions, but only if these organisations are receiving funding from BBSRC. The question of interoperability does not apply to the other two documents [11, 23], because each of them give an overview of a substantial number of policies rather than a single one.

When it comes to policy makers, of the documents analysed in this category, three were penned by the European Commission [11, 13, 14], where document [13] was created with the support of Alma Swan (Convener for Enabling Open Scholarship and director of Key Perspectives Ltd.) and the other two [11, 14] with the support of Neelie Kroes (Vice-President of the European Commission, responsible for the Digital Agenda). The Canadian Heritage Information Network policy [6] was written by the

² BBSRC is just an exemplary for other RCs



institution itself, while the Italian [1] policy and the Finnish [20] roadmap were created by dedicated committees. The JISC policy was prepared by Charles Beagrie Limited [4] and the Indian policy [22] by the Ministry of Science & Technology (Department of Science & technology) within the Government of India.

The document 'Sharing knowledge: open access and preservation in Europe' [13] was designed after a Workshop was held "to explore the state of play and progress within Member States (MS) with respect to Open Access to, and preservation of, scientific research outputs". Therefore, the policies in this document concern all kinds of sectors or groups. Documents [11] and [14] include the Commission's visions and recommendations concerning the open access, preservation and dissemination of scientific information and is meant to be applied across Europe. The Indian policy [22] was designed to apply to all sharable non-sensitive data available either in digital or analogue forms, but generated using public funds by various ministries, departments, subordinate offices, organizations and agencies of the Government of India. The Italian policy [1, 15] applies to the e-government sector, to archival records, as well as private companies and the CHIN checklist [6] to museums in Canada. The Finnish roadmap [20] applies to ministries, research institutions and so on, at a national level, while the JISC policy [4] applies to UK universities and colleges.

Further, it is stated in [13] that "the Commission would like to develop concrete policy recommendations on how to move forward at Member State and European level on access and preservation issues". Preservation is the goal of three more policies [1, 4, 6], one of which, the JISC policy [4], further explains that the aim of the study is to provide a framework for a digital preservation policy, as well connecting to other strategies within universities and colleges in the UK. The Finnish roadmap [20] aims at common open access and data sharing and the Indian policy [22] is designed for sharing and making accessible all non-classified data collected using public funds. Finally, the purpose of the policies [11] and [14] is the dissemination of and open access to scientific publications and research data from publicly funded research and the preservation and re-use of scientific information.

As in the case of SIM4RDM, the question of interoperability does not apply to the [13] policy. The [11, 14] policies, on the contrary, state as their goal that policies should be developed by Member States and implemented by institutions. Three of the other policies [1, 6, 22] can be applied to other organisations at national or international level. The Finnish roadmap [20], however, explains that it may not be transferrable to other countries easily, because of the varying legislation and Government structures.

2.2.3.2 Data policy adherence

In research funding agencies and particularly in the BBSRC policy [5], adherence is monitored through the Final Report assessment procedure, while the SIM4RDM study [23] notes that there are national bodies that provide co-ordination and raise awareness, amongst other things and that these national bodies could possibly draft a Code of Conduct. In the latter case the Commission, along with the European Union, have been suggested as the designated parties for such a code, but there have been no actions to support that as of yet.

In two circumstances there are no mentions of the responsibilities for data curation and preservation. The SIM4RDM study [23], however, describes that "just over one third (36%) of the organisations requiring retention of data designated a specific organization for preservation". As for providing incentives, it is recognized that "of the organisations that responded, 25% created incentives to deposit data". "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". On this matter, the 'National Open Access and Preservation Policies in Europe' document [12] stresses that incentives to encourage researchers to provide open access to their publications are in place in 15 out of 29 responding countries. "In some cases, incentives will be made more explicit when a national open access strategy is in place".



With regard to the persons that inspect whether each of the policies are being adhered to, the Finnish roadmap [20] mentions that there is a large cross-functional collaboration in political, legislative, organisational and infrastructure levels in place, while the JISC policy [4] simply acknowledges the need to carry out surveys and to put audit guidelines in place, in order to assess awareness of policies. In two cases the European Commission is involved: in the first case [12] it directly monitors the progress of the policies, and in the second [13] it is noted that the Commission and the European Research Council should follow up on policies, but that this has yet to happen. In two other cases [15, 22], adherence is monitored by departments or services within the same institutions, which produced the policies.

The Italian policy points out that in each organisation the digital records creator must identify a responsible person for preservation [15], while the JISC policy [4] merely states that preservation strategies should be considered. The Indian policy [22] assigns the curation and preservation responsibilities to each of the data owners and the Finnish roadmap [20] specifies that the curation is to be carried out by research organisations and the preservation by a centralised national unit.

The 'Sharing knowledge: open access and preservation in Europe' document [13] stresses that incentives should be developed for researchers, authors and publishers to make their work Open Access or to change their own practices and norms, thus increasing the amount of Open Access content. The [11] document suggests "putting in place mechanisms enabling and rewarding the sharing of research data", as well as "new reward systems and structures". The Finnish roadmap [20] mentions a merit system for researchers and suggests the creation of supportive/incentive laws. Finally, only two of the policies [4, 20] in this category conduct some sort of risk assessment or risk management.

2.3 DATA REQUIREMENTS, PRESERVATION AND RE-USE

Analysing the data preservation policies has shown that the decision on what kind of data will be preserved is approached in various ways (for example according to digitization process, material format or funding source), and is dependent upon the overall focus of the organization as well as the timeframe foreseen at the time of writing the policy. In this chapter, the preserved data, open access and re-use of data, preservation requirements, time frames of preservation and metadata requirements from the different policies are reviewed against the sectors of memory institutions, universities and research institutions, and funding agencies and policy makers.

2.3.1 Memory institutions

2.3.1.1 Preserved data

Preserved data can be limited on an organizational basis, which means that the data to be preserved includes all the digital objects that one or several named libraries, archives or museums collect or create [19, 26, 27, 28, 30], regardless of the format, origin, the creation time or place of the data, the nature of the medium [27], or whether the material is published or not [28].

Preserved data can be categorized, for example, according to the digitization process. In this case, the material is divided in two groups: 1) born-digital resources, which have been created and managed electronically, and 2) made-digital resources, which have originally been created in non-digital form, but have been subsequently converted to digital form [26, 30, 32]. A third, not so commonly discussed, group is re-made digital resources [32], where the material has originally been created digitally and then it has been transformed in non-digital form, and been subsequently re-digitized.

Preserved records are often master originals (for example in [24]), but the material can also have several surrogates [29, 32], and those can be arranged in different categories by their function, such as for preservation, access and business [32]. The master data to be preserved can be raw data, or data derived from raw or other original format [29]. In this case, special care may be required to respect the rights of the owners of the raw data and the derived data [24].

Besides the content, preserved data can also be classified by their format, such as digitized and originally digital documents, publications, photographs, films, broadcasting archive, private archives,

Date: 2013-12-31



audio recordings, text, multimedia, websites, objects inside websites, digital mapping, databases or metadata records [29, 30, 31].

2.3.1.2 Open access and re-use

Open access and the possibility of re-using data is an important factor in different data policies. The National Archives UK policy [24] requires that the data is openly or freely accessible and reusable. One other aim is to give the users open access as widely as possible, but within the existing financial and legal framework [31].

In the policy of the National Archives of Estonia [27], requirements for access to archival records in the form of regulations, recommendations and guidelines for the various target groups in society can be developed. Original and "user copies" of the digital material can be different and in this case access to the original material will generally not be provided.

According to the policy of the National Library of Australia [28], the technical challenges for ongoing open access from a preservation point of view include: technological hardware and software obsolescence, wide range of file formats, fast growth of digital collections, different preservation approaches, different access policies, complex object structures, relationships and dependencies, and complex preservation planning, processing and action scheduling.

Further, the digital preservation strategy of Archives New Zealand [3] states that the "content should be offered in open, non-proprietary formats"

2.3.1.3 Preservation requirements and time frames

The policy of the National Archives of Australia [26] states that the preservation file formats used should be open and that there should be no patents, intellectual property rights or other rights attach to the format. Also, migrations to more recent file formats can be done, if needed [30]. On the other hand, the format restrictions can be defined at a very general level [28]: "While there are no predetermined restrictions on the formats that will be accepted into the Library's repository, a decision about whether or not to accept the material will need to take into account the current level of support that can be provided for the known formats contained in the collection or collection item."

One wider example is found in the policy of the National Archives UK [24], where the policy principles provide specific details in the following areas:

- 1. Vision preservation activities are embedded into all activities.
- 2. Access and authenticity preservation is crucial in allowing current and future access to records and ensuring the authenticity of records.
- 3. Value of records in terms of information, intrinsic/material, or cultural
- 4. Risk-based approach preservation of risk assessment for physical items which is reviewed regularly and for electronic records, which are individually assessed on a regular basis and inform the digital preservation planning process
- 5. Conservation preserving record with minimal impact on its authenticity. Preventive and conservation treatments are used
- 6. External role leads nationally and internationally on archival preservation in terms of expertise, implementation and dissemination of standards and best practice.
- 7. Digital preservation preservation of electronic records focuses on long-term access of the content and associated value rather than the technical elements of the record (due to obsolescence). There two levels of preservation:
 - a. passive preservation providing secure storage and integrity of the record;
 - b. active preservation requiring the migration of records to new technologies. The original will always be stored along with any subsequent records

Many of the data policy related studies neither mention nor try to analyse – or even predict – the preservation timescale of the data. This factor may not be covered at all, or may be covered by the



general idea of long-term [31], or with phrases such as "the longest possible period" [27] or "as far as possible" [24]. But instead of preserving data forever without any real estimate, it is possible to make a plan or develop guidelines about what data should be preserved and for how long. For example in the policy of National Museum of Australia [30], a tough description is given: "A mandatory retention period should be identified for all digital records, to ensure that digital files are only kept as long as they are of value to the Museum. The retention period will be reviewed periodically, and assessed prior to the disposal of a digital file."

2.3.1.4 Metadata requirements

According to policy of National Museum of Australia [30], metadata is required for "ongoing preservation, discovery, access and rights management of digital assets is captured". The policy document defines metadata "of both the cataloguing information associated with the content of the digital file and technical details of the file, as well as details of its relationship with other files".

In the policy of National Library of Australia [28], format recommendations towards international standard PREMIS [33] can be found. Also, some national standards are used, for example AGLS (Australian standard: AS 5044:2010) [25] is expressed in the policy of National Museum of Australia [30].

For surrogates, the copyright must be observed, and the metadata must contain a detailed migration history [24]. Periodic review of preferred digital metadata standards is also stated in the policy of National Library of Australia [28].

2.3.2 Universities and research institutions

2.3.2.1 Preserved data

The data to be preserved can be raw data, and it can be categorized by the research process [21]:

- (1) Data continuously acquired, which is not related to funded projects,
- (2) Data acquired by instruments and sensors, which usually is related to funded projects, or
- (3) Data acquired by other means or derived from samples analysis.

2.3.2.2 Open access and reuse

DISC-UK [8] encourages that data be openly or freely accessible and reusable. According to Winkler-Nees [39]: "the willingness to provide access to data for scientific re-use should become standard in science and research". Data processing should be considered when choosing the manner of open access, for example finding "the most suitable way for the submission of data, including calibrating and formats, considering the different cases that could appear in such a complex situation", as noted in the Spanish policy [21]. There is also a need to "develop and implement appropriate infrastructures and tools" [39].

Open access can also be restricted for certain groups or purposes. The policy by Sapienza University of Rome [34] includes three types of access level, depending on the data: internal access to the specific research team; open access to the organization community and general open access. In some countries the possibility of true open access is fundamentally restricted by legal issues. A strong statement can be found about Swedish policies by Slavnic [36]: "Preservation of data, re-use and access is hardly possible due to Personuppgiftslag (The Personal Data Act) that says that it is illegal to use research data other than in specific projects for which the data were collected." This said Swedish National Data Service (SND) is working on making data available. The Swedish Research Council has also a data publication plan [37].

2.3.2.3 Preservation requirements and time frames

In general, preservation requirements and time frames are not very well discussed in the policies by the universities and research organizations. However, Sapienza policy [34] defines very specific guidelines for the file formats, where for example file formats TIFF, XML, PDF and JPEG are included. In the Spanish policy [21], the preservation period for the research data is from four years



onwards or for a maximum of four years, depending on whether the material is freely available or not, respectively.

2.3.2.4 Metadata requirements

A set of minimum metadata elements are defined in the Spanish policy [21] and recommended in study [2]. Standards for description and indexing according to the specific scientific domains and a format recommendation for usage of the international PREMIS standard [33] is mentioned in Sapienza policy [34]. The Swedish National Data Service's metadata requirements contain a web form for personal, study, funding and dataset information³.

2.3.3 Research funding agencies and policy makers

2.3.3.1 Preserved data

The data in question might be diverse, and for example in BBSRC policy [5], attention is specifically paid to data arising from high volume experimentation, low throughput data arising from long time series or cumulative approaches, and models generated using systems approaches. The data to be preserved can also be restricted to data provided by a certain funder [34], or is concerned with data provided by public funding in general [20].

2.3.3.2 Open access and reuse

Some public funding rules may require that the data is openly or freely accessible and reusable [12]. In some cases, the requirement for data availability is strictly defined in the funding policy [5]: "All applications seeking research grant funding from BBSRC must submit a statement on data sharing. This should include concise plans for data management and sharing as part of research grant proposal or provide explicit reasons why data sharing is not possible or appropriate." and "BBSRC recognises that different approaches to data sharing will be required in different situations and considers that it is most appropriate for researchers to determine their own strategies for data sharing and outline these within their research grant proposal(s). Applicants should consider where, how, and to whom their data should be made available. "According to 'Sharing knowledge: open access and preservation in Europe' document [13], "experts from the Member States where national-level or institutional-level policies have been adopted reported that they are successful in increasing the amount of material openly available and in raising awareness of Open Access amongst authors".

The Finnish roadmap [20] notes that the research data should be extensively available for research, education and policy-making. To solve the questions arising along the aim of open access, one solution is to develop supporting open data services [15]. Also, some additional notes can be found in SIM4RDM document [23]: "Organisations facilitate data re-use mainly by creating data repositories" and "The organisations also mentioned websites for accessing and re-use of data. Other solutions that make re-use of data feasible are lists of available datasets in-house or on the intranet".

2.3.3.3 Preservation requirements and time scale

For digital objects, the definition of suitable file formats is not widely considered in data policies. A checklist regarding support, openness and proprietary formats, the possibility to include metadata, compression, versions of software, data loss and so on are given in the document by Canadian Heritage Information Network (CHIN) [6], and it also includes questions regarding vendor, environmental fluctuations and conditions, handling conditions, etc. The Beagrie study funded by JISC [4] provides guidance on the preservation strategies which can be adopted – either lifecycle-based or following the OAIS reference model [17]. In Italian documents [1] and [15] very specific guidelines for file formats are defined, where file formats PDF, PDF/A, TIFF, JPEG, OOXML, ODF, XML, TXT and email messages are discussed. According to 'National open access and preservation policies in

³ http://snd.gu.se/en/deposit-data/webform

Grant Agreement 269977



Europe' document [12], "not all countries have explicit policies or arrangements for research data, but some funding bodies are encouraging making research data openly accessible".

Some of the studied policies either do not consider persistent identifiers, or consider them at a very general level. For example, in the Finnish roadmap [20], a centralized persistent identifier service is suggested. URL and DOI⁴ formats are given as examples in CHIN [6] and SIM4RDM [23] studies. According to SIM4RDM document, publishers use a DOI as a persistent identifier for the citation of data, and also national bodies should endeavour to standardise conventions for data citation.

It is stated in SIM4RDM document [23]: "Several factors contribute to determine how long data must be preserved. Preservation for more than five years is determined by the function of the institution". BBSRC recommends [5] that data should be maintained for a period of ten years after the completion of the project. However, in the Finnish National roadmap for the utilization of electronic data in research [20] the preservation timescale is estimated as "dozens or hundreds of years". The required preservation time is then dependent, for example, on the data (as noted in paper [15]) and the institution. Therefore, all relevant detailed information about the timescale should be included in the preservation strategies sections of the policy document as is recommended in the Beagrie study [4] or in the data preservation plan.

2.3.3.4 Metadata requirements

The BBSRC policy [5] states that it is "needed to provide a secondary user with any necessary details on the origin or manipulation of the data in order to prevent any misuse, misinterpretation or confusion." Also, the SIM4RDM document [23] states that "Where discoverability and reuse are concerned, sufficient metadata should be included that are openly available to others".

Recommendations for the use of the international Dublin Core standard [9] can be found in 'Sharing knowledge: open access and preservation in Europe' document [13]. Also, some national standards are used: for example, UNI 11386:2010 SInCRO (Italian standard) [10] is found in policy [1], where the requirements are specific with reference to the definition of metadata related to authorship, dates, subjects, etc. for public administration. In the CHIN's checklist [6], there are questions regarding standards, interoperability, aggregation, authenticity indicator, and so on although concrete suggestions are not provided.

2.4 OWNERSHIP AND RESPONSIBILITIES

Responsibilities associated with the data ownership, and with reference to the nature and the role of the repositories, have consequences for the type of policy in place, for its framework and for the details of the controls and the rules established and their regular update. We analyse these aspects here, with reference to major standards and recommendations like ISO 16363 [18] and TRAC [7] and to national legislation and regulations. We also factor in the stakeholders' functions: memory institutions, universities and research institutions, funding agencies and policy makers.

2.4.1 The memory institutions

The memory institutions express adequate awareness for defining and implementing policies on a common basis and with a strategic perspective, even though the responsibilities are not detailed; they generally refer to an organization chart and internal administrative hierarchy.

In only one case are the responsibilities well detailed and differentiated: the National Archives UK [24] has assigned the definition of policies for risk assessment to the director, while the responsibility for digital preservation, for updating and disseminating the related policy is confined to the department heads and policy implementation is considered an obligation for all staff.

In the policies of National Archives of Estonia [27] and Norwegian Ministry of Culture [31] a standardized approach is mentioned with reference to the definition of responsibilities and with an implicit recognition of the roles defined in the ISO 16363 [18] or TRAC [7] specifications. In Norway a specific council for digitization and digital preservation – with key players and a specific mandate

⁴ http://www.doi.org/

Grant Agreement 269977



established by the Ministry of government administration and reform - has the formal task of providing recommendations to the Ministry of Culture in coordination with the Norwegian standard council.

Responsibilities for internal policies are described with reference to the activities involved by the National Library of Australia [28], with specific attention to access, training, technological control and infrastructure. The National Archives of Australia [26] also pays attention to cooperation at a national level. Internal cooperation for building policies for digital preservation is considered crucial by the National Museum of Australia [30] – and even more so for implementing them.

For all the memory institutions studied, the framework for licenses is based on national legislation. In only one case [31] was a specific project (Bokhylla.no) dedicated to defining a model for a collective licence arrangement, which would be extendible to other institutions and universities.

2.4.2 Universities and research institutions

Five institutions have been identified in this analysis. The results cannot be considered exhaustive but offer useful information and indicate common trends. All the institutions recommend public access, but cannot and do not place such obligations on researchers. An effort at standardization, at least at national level, has been carried out by the consortium of German Universities and research centres [2] via license agreements. Ownership usually rests with researchers. Only Sapienza Digital Library SDL [34] explicitly declares in its mandate that data and records are public and pertain to the University, with the exception of digital resources donated to the SDL. Attention is mainly concentrated on copyright and access policies. No effort seems to be dedicated to digital preservation, except for the case of SDL where the collection policies do try to include rules and workflows for digital preservation. In this case, the policy carefully defines all the responsibilities involved:

- in the preparation and management of data and digital objects (individual department, University Library or the research centre for digital humanities Digilab);
- in the ingest processes, University library and Digilab (two different flows for different resources), according to the rules and the strategy defined by the scientific committee and the management committee of SDL;
- in the storage and preservation environment (at the moment outsourced to a public consortium, CINECA) to be implemented according to a very detailed policy based on ISO 16363 [18] and national legislation.

Mechanisms for updating policies are not normally made explicit, again with the exception of the Sapienza Digital Library (Italy) [34] which has assigned this task to the SDL scientific committee.

2.4.3 Research funding agencies and policy makers

These two groups of cases, even if characterized by a different mandate, can be analysed together and the results of the desktop research (eight institutions and projects) confirm common trends. In this case differentiation is even greater than in the other two categories because of the specific jurisdictions in place. In all cases, the main recommendation is always open access and the need for a more standardized approach for licenses is always considered crucial, although not currently supported with specific and dedicated projects. The European studies and reports focus strongly on the need for planning standardized approaches for open access and open copyright systems and of defining clear and well stated responsibilities: see [23], which recognizes the relevance of covering "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" and further recognizes that "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".



The main effort is, of course, invested by national bodies even if the goals vary considerably. The most detailed and comprehensive policies are those approved in 2012 by the Government of India [22] and the legislation under approval in Italy (under the coordination of Agenzia per l'Italia digitale) with specific reference to digital preservation policies.

The Indian public policy [22] for national data sharing and accessibility is very detailed with respect to responsibilities, despite the fact that the policy is not dedicated to digital preservation. With reference to security procedures, the guidelines imply that computer services department within each organization serve the central functions of coordination and monitoring of policy. In this way the policy provides a national coordination on the need to include standards and define metadata. Creators are responsible for publishing their data, and defining and regularly updating their policies.

The Italian legislation⁵ is very detailed and specific with regards to responsibility for digital preservation and also with respect to updating mechanisms. The digital records creator must identify a responsible person for preservation, and each repository can be certified only if a similar role responsible for preservation is in place. This responsibility must be defined by the managers for IT departments and the chief of records management (for public sector bodies). Regular updates are foreseen – every three years for the national regulations and technical guidelines – while the manual for digital preservation internal to each organization must be updated when necessary, accordingly with changes of technologies, legislation and regulations.

The Finnish roadmap [20] for research infrastructure is also useful for identifying responsibilities, if not for preservation, then at least for the data management life cycle, which is the first requirement for future persistence at national and institutional level. They include: common rules and common services defined by the ministries; portal services, metadata tools, processing services, schemas, thematic vocabularies and ontologies under the responsibility of research organizations; production services for data, interfaces, mechanisms to transfer the data, descriptions of the processes to provide and maintain data under the responsibility of research organizations individually and finally the task of providing data and related metadata on the part of researchers and projects.

3 THE SURVEY

In addition to the desktop research described in Section 2, a web-based questionnaire was conducted to obtain a view of the kinds of policies and governance structures are in use in different organizations. The questionnaire was sent to several international mailing lists together with an invitation to participate an APARSEN questionnaire on Digital Rights and DRM⁶, and it was also advertised in some national bodies. The questionnaire was open from September 5th to October 14th, 2013. The questionnaire is provided in Annex 6.2 for reference.

A total of 27 respondents participated in the survey from ten countries (see Figure 1). Most of the respondents were from Europe with only two from outside Europe. One third of respondents were from the APARSEN consortium members and with two thirds outside of APARSEN. Since the questionnaire was sent to several mailing lists for which we do not know the exact number of recipients, the response rate cannot be specified. However, since we received only 27 respondents, limited, although interesting conclusions can be drawn from the results.

⁵ to be formally approved in December 2013

⁶ See Deliverable D31.1 Report on DRM preservation





Figure 1: Answers by countries (n=27)

In the survey, 70% of the answers were from four countries (Finland, Italy, The Netherlands, and United Kingdom). In Section 3.2 we describe the results from these countries in more detail.

Most of the respondents were from universities (34%), research data archives (19%), national libraries/archives (13%) and research institutes (13%) (see Figure 2). Figure 3 summarises the respondents' role in data.



Figure 2: Organization profile



Figure 3: Role in relation to data

3.1 SURVEY RESULTS

Most of the answers were related to internal policies of an organization (59%) or common national policies (56%) (see Figure 4).





Figure 4: Sector to which data policy is related in general

Almost all policies covered in this survey were related to digital preservation (93%). Data sharing was also well covered (78%), as well as open or authorized access to data, 67% and 63% respectively. Figure 5 summarizes the purposes of the policies. More than half of the policies covered at least three purposes and 26% of the policies covered all selected purposes. Only a few policies covered other purposes, in which cases, the purpose of the policy was mainly for re-using the data. Most of the policies covered raw (78%) or published data (74%), and 62% covered them both. The owners of the data vary considerably between policies. In many cases the depositor or funder is the owner of the data, but it was mentioned in several answers that the owner depends on the funding scheme or how the data is created. Thus the ownership is in many cases decided case by case.





Although many policies were related to digital preservation, only a few specify for how long the data should be preserved. In more than half of the cases, the preservation time is undecided or not known. Similarly, the timeframe for providing access to data depends upon the data itself. For example, none of these policies specify a moratorium time. Only three policies specify that data is immediately accessible and one policy specifies that data will never be openly available.

As shown in Figure 6, which summarises the license types recommended in policies, 56% of policies recommend and/or require some license for the data (open, collection of various types, international, or national license), but in most cases the license is not specified, or any kind of license is accepted. Only a few policies allow only one type of license. Furthermore, many policies prefer open access as far as possible, although it is widely acknowledged that this is not always possible.





Figure 6: Licenses recommended in the policy

Metadata of high quality is required for digital preservation or data sharing to be successful. Almost half of the policies require descriptive metadata to be attached to the preserved data (see Figure 7). Provenance, technical and rights metadata is required or recommended in roughly half of the policies.





The creation and management of metadata is not the only requirement to preserve the related digital objects. For example, about 50% of policies require persistent identifiers for data (see Figure 8). The OAIS model [17] is used in about 50% of the systems related to these data policies. File formats and categories are also specified and/or required in many policies, but only very few policies consider file size limits. Some policies, however, are high-level and therefore do not consider technical details at all. Other technical requirements are typically discipline-dependent issues, important in specific cases but perhaps not in general data policies. Technical requirements, as well as metadata requirements, are a key issue in enabling interoperability between repositories.





Figure 8: Specific requirements to the preservation activities

In this survey, the funding behind the data policies is based on short-term project-based funding in more than a third of the cases (see Figure 9). Such funding models cannot guarantee long-term digital preservation. However, in some cases basic activities are funded on lump-sum basis, but in some specific activities project-based funding is used.



Figure 9: Funding model

Whether one's governance structure would be relevant to other organizations is, according to the responses received, something that has not been considered by many organizations, or the governance structure is mainly seen as being too complex to benefit others. The applicability of data management mechanisms to other organizations is seen in a much more positive light: many respondents reported that cooperation in this field was already taking place, mainly between organizations with similar activities in regard to data archives, metadata standards and IT solutions that enabling sharing (e.g., iRODS⁷).

Most of the organizations have a steering group responsible for policy management (see Figure 10). Not all organizations have a single dedicated body for policy management, but this task is distributed among the organizations.

⁷ https://www.irods.org/

Grant Agreement 269977





Figure 10: Organisation's governance structure in relation to data policy management

Interoperability issues in data management are important to ensure that the preserved data are reachable and useful to others and thus enabling cross-disciplinary research. Many respondents of the survey highlight the importance of interoperability with respect to data policies, but in practice it is not realized. According to the responses the most efficient way to make organizations think about interoperability when they formulate data policies appears to be to have a funder require it. Some respondents report some degree of cooperation at the time of creating data policies, but mostly with organizations with similar function or structures. Generally, the interoperability question is seen as an issue which is not yet being addressed and is not reflected in existing policies. One positive example of effort of taking interoperability to the level of implementation seems to be Research Data Alliance's (RDA)⁸ activities in sharing data policies, as mentioned in some answers.

According to the survey, collaboration with other institutions does exist, either through on both official agreements defining the scope of collaboration (for instance, in relation to digital preservation, common services and development work), or less formal collaborations, based rather on the exchange of information, for example through working groups. Many respondents define collaboration agreements through suitable projects or networks, such as CESSDA⁹, CLARIN¹⁰ and EUDAT¹¹. In terms of interoperability, participation in collaborative projects can serve as a low threshold activity especially for organizations without strong experience of collaboration on data policy issues.

3.2 RESULTS IN SELECTED COUNTRIES

70% of the answers to this questionnaire were from four countries: Finland, Italy, The Netherlands, and United Kingdom. It should be noted that there were only 4–5 answers per country, and thus the following analysis should be treated as indicatively only, and no far reaching conclusions should be drawn.

In the policies from the UK, digital preservation, data sharing, open access and authorized access is taken into account in all policies (see Figure 11). The same is true also in the Netherlands, except that authorized access is the lowest among all the answers. In Italy, access to data (either open or authorized) seems to be taken into account less than in other countries, although data preparation is considered to be more important.

⁸ http://rd-alliance.org/

⁹ http://www.cessda.org/

¹⁰ http://www.clarin.eu/

¹¹ http://www.eudat.eu/

Grant Agreement 269977

Date: 2013-12-31 D35. Project: APARSEN Doc. Identifier: APARSEN-REP-D35_1-01-1_0





Figure 11: Purpose of the data policy in selected countries (n=19)

Regardless of the country, raw and published data are the major subjects of the data considered in the policy, as well as in the survey as a whole (see Figure 12). However in Italy, published data has quite a low emphasis in the policies considered in this survey. Comparing with the purpose of the data, one can see that the policies in Italy are more about data preparation than about sharing and access.



Figure 12: Type of the data in selected countries (n=19)

In designing and writing data policy, in Finland the ministry of Education and Culture collaborates strongly with national institutes and dedicated task forces, whereas in Italy dedicated task forces have a slightly higher impact. However, in the Netherlands and UK national institutes play a bigger role. Data ownership rests mainly with the depositor in Finland and Italy, whereas in the Netherlands the funder seems to have a bigger role (Figure 13). In the UK, the ownership of deposited data depends mainly on how the data are generated or to which category they belong.





Licensing policies vary a lot between these countries. In Finland, copyright protected materials are dominant, whereas in Italy and the Netherlands open licenses or some uniform collection of licenses are preferred. Persistent identifiers are required in Italy in 100% of the policies surveyed, compared with only 20% of cases in Finland. Other specific technical requirements follow pretty much the average in the whole survey (see Figure 14).





4 RECOMMENDATIONS AND CONCLUSIONS

Following the same structure as in Chapter 2, we present recommendations and conclusions for the organisations considered using the same four categories.

4.1 GOVERNANCE AND FUNDING

Our desktop research provides recommendations for future work from the Beagrie study [4], which states that defining a digital preservation policy, as well as the implementation of such a policy, is considered as core business activity and should thus be embedded into an organization's overall policies, with clear mappings to other key policies within the organization. It also states that financial sustainability should be defined clearly as part of the overall organizational financial plan.

Based on the results of the survey data policy is most often related to digital preservation. However, funding is usually short-term and project-based, which creates risk for preservation. To protect data persistence, solutions of more permanent funding are required.



Recommendation 1: Sustainability of funding for digital preservation and access should be assessed. In the case of research funder's policy, is there separate funding available, for example, in research grants for data management to follow the policy that should be communicated clearly for policy users.

Recommendation 2: When defining a digital preservation policy, as well as the implementation of such a policy, it should be considered as a core business activity and should thus be embedded into an organization's overall policies, with clear mappings to other key policies within the organization.

4.2 DATA POLICY AND DATA POLICY ADHERENCE

Most of the memory institutions write their policies by themselves and for themselves. However, half of the memory organizations consider that their policies could be applicable to other similar institutes, nationally or internationally. In the research field it is more common that policies are written by several contributing partners or dedicated committees. Most of the policy makers' policies can be applied to other organisations at national or international level. To increase interoperability between organizations, we recommend that policies should be written through close collaboration between similar institutes, although some of the policy details may be organizationally specific. This conclusion was borne out by the desktop research and the survey results. The possibilities of nationally – or even internationally – unified policies should be considered, even though this may not be easy (especially in the latter case because of varying legislation and government structure). Also, in national and international cases, close collaboration with the governmental level is recommended to find the right compromise between the need for open access and reusability, and legal frameworks.

Recommendation 3: Cooperation with similar organizations and users should be stated within a policy on digital preservation and access. Describing the level of cooperation and possible agreements are noteworthy. The applicability of policy to other organizations and possible benchmarking activities for future purposes is recommended.

Recommendation 4: A policy should be made available on the web and be easily retrieved.

Usually, the purpose of data policies is preservation in memory organizations, while policies of universities and research institutes highlight open access and sharing. Both kinds of policies (preservation or access) can be found from the documents created by the policy makers. However, to enable long-term accessibility, it is recommended that both parts should be well-covered. In the survey, preservation was considered in almost all responses, and open access in 2/3 of responses.

Recommendation 5: A policy should include the reason for the policy and the target group.

Recommendation 6: A policy should encourage open access to data. If needed, the policy should recommend embargo periods, solutions for handling privacy concerns and so on.

In memory or research organizations, procedures for adhering to policies are not covered very well. In memory organizations, monitoring adherence to policy is done internally, whereas in the research field, steering groups and strategic boards are used. Funding agencies and policy makers lean toward separate procedures or code of conducts. It is recommended that adherence to policy be monitored by steering groups and through close collaboration between organizations. Also, survey work may add value in identifying bottlenecks in procedures.

Recommendation 7: The mechanisms for updating the policies should be explicit.

Some of the organizations provide incentives for actions which support preservation, access and reusability, but this uncommon. Policy makers suggest that incentives should be developed for



researchers, authors and publishers to make their work openly accessible. Similarly, supportive incentives for actions that support preservation and reusability should be considered.

Recommendation 8: Incentives to participate should be provided. The skills required from users in order to be able to follow the policy should be identified, and the availability of any support services should be described in the policy.

Risk assessment or management procedures are not very well covered in the policies surveyed. They were apparent in half of the policies used by memory organizations, but not included at all in policies in the research field. For securing long-term preservation and availability of the data, it is recommended that risk assessment and management procedures be covered in the data policy documents.

Recommendation 9: A policy should include risk assessment with contingency planning and an exit strategy.

4.3 DATA REQUIREMENTS, PRESERVATION AND RE-USE

Categorisations of preserved data are typically considered in each institute or funding agency. The data that are preserved are usually the ones considered as master copies, although surrogates may exist.

The data processing steps, infrastructure, and provided tools and services should all be created to support preservation, open access and reuse. Also, issues related to vendor, environmental fluctuations and conditions, handling conditions and so on should be considered in the policies.

Required preservation time is dependent on the data and the institution. Timescales should be included in the policy documents or in the data preservation plan. One option is to review the preservation time periodically and separately for each data set.

Support, openness, patents, the possibility of including metadata, compression, versions of software, data loss, and so on should all be considered when choosing suitable file formats. Commonly used open file formats increase interoperability as well as common agreed metadata standards that document the digital objects.

In the survey, more than half of the responses required or recommended common metadata formats. We recommend the use of international standards for metadata, such as PREMIS [33] or Dublin Core [9]. If international standards are not suitable, some self-generated profile based on an international or national standard should be used. Common specifications based on international standards increase interoperability and comprehensibility of data. Different tools for providing metadata in a unified form should be developed. If international or national standards are not suitable and a new metadata format is needed, it is worthwhile considering creating it through a standardization process.

Persistent identifiers are not very well covered in the studied policies, although in the survey about 50% of the respondents used them. To increase interoperability and "referenceability", the use of global persistent identifiers for both metadata sections and objects is recommended.

Recommendation 10: A policy should include the type of data addressed, from published papers or results funded under a particular programme, and so on. A digital preservation policy should always include some elements dedicated to digital preservation (file formats, metadata requirements, etc.) and to the appraisal/retention rules, both with reference to international standards and to national legislation and rules. The policy should give guidance on how long data needs to be preserved.

Recommendation 11: A policy should include specification of standards and interoperability. Any requirements for compliance with standards should be clearly defined, even if the list of requisite standards varies according to the specific nature of the institution.



4.4 OWNERSHIP AND RESPONSIBILITIES

In memory organizations ownership policies are usually based on institutional mandates and relate to national legislation, while in the research field the ownership usually resides with the researchers. It is recommended that open and standard licenses (e.g., Creative Commons licenses¹²) should be considered, which clearly state how the data can be used or shared. This was also evident in the survey responses.

While in memory organizations responsibilities for data curation and preservation are an internal matter, in the research field this is laid at the door of a strategic board or a scientific committee, or sometimes even an external service. In other cases, the responsibilities and roles are not very well defined in the policies. Responsibilities and roles should be recognized according to international standards (for example ISO 16363 [18] and TRAC [7]) and maintained by a steering group.

It is also worthwhile considering if the whole data management life cycle should include common rules, services, tools, vocabularies, interfaces, and descriptions of the processes and so on.

Recommendation 12: A policy should specify who to contact to actually preserve and share data, and who takes care that the policy is implemented. Responsibilities for implementing the policy should be based on the organization chart of the institution according to the internal administrative roles and hierarchy. The policy should make a clear distinction between various areas of responsibilities: risk assessment, updating and disseminating the policy, and implementing and respecting the policy.

Recommendation 13: The license systems should be standardized and simplified, or at least clearly expressed. If possible, the policy should give recommendations on licenses to use, as well as possible reuse restrictions. Further, workflows for acquiring digital resources and for identifying obligations and rights for users must be declared and clearly described.

5 REFERENCES

- [1] Agenzia per l'Italia digitale. <u>Gestione documentale</u>, Italy, 2013.
- [2] Alliance of Science Organizations in Germany. <u>Priority Initiative 'Digital Information'</u>, Germany.
- [3] Archives New Zealand and National Library of New Zealand. Digital Preservation Strategy, New Zealand.
- [4] Neil Beagrie, Najla Semple, Peter Williams, and Richard Wright. <u>Digital Preservation Policies Study</u>, JISC, United Kingdom, 2008.
- [5] Biotechnology and Biological Sciences Research Council (BBSRC). <u>BBSRC Data Sharing Policy: version 1.1 (June 2010 update)</u>, United Kingdom, 2007.
- [6] Canadian Heritage Information Network. <u>Checklist for Creating a Preservation Policy</u>, Canada.
- [7] The Center for Research Libraries (CLR), <u>Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist</u>, 2007.
- [8] Data Information Specialists Committee UK (DISC-UK). <u>Policy-making for Research Data in Repositories: A</u> <u>Guide</u>, United Kingdom, 2009.
- [9] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1, 2010.
- [10] Ente Nazionale Italiano di Unificazione. <u>Supporto all'Interoperabilità nella Conservazione e nel Recupero degli</u> <u>Oggetti digitali (SInCRO). UNI 11386:2010, 2010.</u>
- [11] European Commission. <u>Commission recommendation on Open Access to and Preservation of Scientific Data</u>, 2012.
- [12] European Commission. <u>National open access and preservation policies in Europe</u>. Analysis of a questionnaire to the European Research Area Committee, 2011.
- [13] European Commission. <u>Sharing knowledge: open access and preservation in Europe</u>. Conclusions of a strategic workshop, 2010.
- [14] European Commission. <u>Towards better access to scientific information: Boosting the benefits of public investment in</u> research. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2012.
- [15] Mariella Guercio. <u>The Italian case: legal framework and good practices for digital preservation</u>, 2012.
- [16] The International Foundation for Information Technology. <u>Taxonomy of key policy types or categorization</u>, 2013.

¹² http://creativecommons.org/

Grant Agreement 269977



- [17] International Organization for Standardization. <u>Open Archival Information System -- Reference Model. ISO</u> <u>14721:2012</u>, 2012.
- [18] International Organization for Standardization. <u>Space data and information transfer systems -- Audit and certification</u> of trustworthy digital repositories. ISO 16363:2012, 2012.
- [19] Library and Archives Canada. Preservation Policy, Canada, 2001.
- [20] Ministry of Education and Culture. <u>Tieto käyttöön. Tiekartta tutkimuksen sähköisten tietoaineistojen</u> <u>hyödyntämiseksi</u>, Finland, 2011.
- [21] Ministry of Education and Science, and the Spanish Geological and Mining Institute. <u>Protocolo de Remisión,</u> <u>Almacenamiento y Difusión de Datos Antárticos en España</u>, Spain, 2007.
- [22] Ministry of Science and Technology. National Data Sharing and Accessibility Policy (NDSAP), India, 2012.
- [23] Wilma Mossink and Magchiel Bijsterbosch. European Landscape Study of Research Data Management (SIM4RDM), 2013.
- [24] The National Archives. <u>Preservation policy</u>, United Kingdom, 2009.
- [25] National Archives of Australia. AGLS Metadata Standard. AS 5044:2010, 2010.
- [26] National Archives of Australia. <u>Digital Preservation Policy</u>, Australia, 2011.
- [27] The National Archives of Estonia. Preservation Policy of the National Archives of Estonia, Estonia, 2007.
- [28] National Library of Australia. Digital Preservation Policy 4th Edition, Australia, 2013.
- [29] The National Library of Wales. Digital Preservation Policy and Strategy, United Kingdom, 2008.
- [30] National Museum of Australia. <u>Digital Preservation and Digitisation Policy</u>, Australia, 2012.
- [31] Norwegian Ministry of Culture. <u>National Strategy for Digital Preservation and Dissemination of Cultural Heritage</u>, Norway, 2009.
- [32] Parliamentary Archives, <u>A Digital Preservation Policy For Parliament</u>, United Kingdom, 2009.
- [33] PREMIS Editorial Committee. <u>PREMIS Data Dictionary for Preservation Metadata, Version 2.2</u>, 2012.
- [34] Sapienza University of Rome. Data management policy for Sapienza Digital Library, 2013 (to be published).
- [35] Madeline Sheldon. Analysis of Current Digital Preservation Policies. Archives, Libraries, And Museums, 2013.
- [36] Zoran Slavnic. <u>Towards Qualitative Data Preservation and Re-Use—Policy Trends and Academic Controversies in</u> <u>UK and Sweden, Forum Qualitative Sozialforschung / Forum: Qualitative Sozial Research</u>, 14(2), Art. 10, 2013.
- [37] Swedish National Data Service. <u>A quick guide to the Swedish Research Councils (VR) data publication plan</u>, Sweden, 2013.
- [38] Swedish Research Council. <u>The Swedish Research Council's Guide to Infrastructure: Recommendations on long-term</u> research infrastructures by the research councils and VINNOVA. Vetenskapsrådet rapportserie 5:2008, Sweden, 2008.
- [39] Stefan Winkler-Nees. German Policies and Approaches to Research Data Infrastructures, ANDS Research Data Infrastructure Workshp, Prato, April 11-13, 2011.



6 ANNEXES

6.1 LEGAL REQUIREMENTS

6.1.1 Memory institutions

- The National Library of Wales: Digital preservation Policy and Strategy The NLW is subject to legislation which governs its activities. Relevant legislation includes The Legal Deposit Libraries Act of 2003, The Freedom of Information Act, 2000 and The Data Protection Act, 1998.
- Preservation policy of the National Archives of Canada There are three Canadian Acts cited. The National Archives of Canada Act, the Copyright Act and the Access to Information Act.
- Digital preservation policy of the National Archives of Australia The Australian Archives Act 1983 is mentioned as a frame for this policy.
- Digital preservation strategy for the Archives of New Zealand and the National Library of New Zealand The Public Records Act 2005 is mentioned.
- National Strategy for Digital Preservation and Dissemination of Cultural Heritage (Norway) Public access to digital documents may be general or restricted, for a number of legal reasons such as copyright protection, privacy, etc. Furthermore, all use of deposited material must comply with copyright law.
- The National Archives (TNA) (UK) TNA follow standard and best practice within the archival sector and have influence within this sector, while effectively setting out the requirements to government departments.

6.1.2 Universities and research institutions

- German policies and approaches to research data infrastructures Fundamental legal frameworks (for example rights relating to data) will be defined.
- Overview of the Swedish status The Personal Data Act is mentioned, which expresses that it is illegal to use research data other than in specific projects for which the data were collected.
- Data management policy for Sapienza Digital Library The national legislation implies specific rules for copyright and for preservation, which have been approved by the Agenzia per l'Italia digitale.

6.1.3 Research funding agencies and policy makers

- National Open Access and Preservation Policies in Europe There are references to certain laws concerning several countries throughout the document, which are too many to be cited here.
- European landscape. Study of Research Data Management (SIM4RDM) In general, it is stated that "legal issues, access policies and long-term archiving are elements that must be evaluated".
- Sharing knowledge: open access and preservation in Europe (Brussels Workshop) "Exploration of copyright laws in EU states with a view to recommending modification or creating a new law on academic research outputs ... to support or permit Open Access" is required. Also noted are certain more specific legal issues, like the need to "clarify legal issues related to linking sharing and re-using Open Access content". "At European level a most significant contribution could be made if it could be ensured that copyright law cannot be overridden by contract law".
- EC recommendation on Open Access to and Preservation of Scientific Data There are legal requirements (that are not specifically illustrated) that concern every Member State separately.



- The Italian case: legal framework and good practices for digital preservation General legislation (Codice dell'amministrazione digitale) approved in 2005 and updated in 2010.
- National Data Sharing and Accessibility Policy (NDSAP) India The policy is in itself a national layer referred to the legal framework of the country.
- National policy in Finland Updating the law is considered to serve the possibility to create and use the research data in a common way.

6.2 QUESTIONNAIRE



Data Policies and Governance

<u>APARSEN</u> is devising a methodology for implementing governance structures and data policies based on current best practice across various repositories.

This questionnaire is aimed at providers, curators, users and researchers of any type of data, who are aware of or responsible for data policies, governance structures and interoperability

issues within their organisation.

Records of Science in Europe Network

Please take some time to complete this questionnaire and share your views on data policies and governance with us. The questionnaire takes about 20-30 minutes to complete and is split into three sections: (1) Current data policies, (2) Governance structure, and (3) Interoperability. Most of the questions are mandatory so please answer n/a (not applicable) if you don't know the answer to the question.

The results of this questionnaire will be published as a public APARSEN deliverable at the beginning of 2014. All respondents will be treated anonymously.

We thank you for your participation.

1. Personal details and contact information



I would like to get notifications for other data related questionnaires

- O Yes
- O No

CURRENT DATA POLICIES

Data management is increasingly becoming policy-driven. In this section, we inventory look at what kind of data policies related to research data are in place today, and how those these are applied in your organisation. This section is split into two three pages.



2. Are there any data policies for preserving research data relevant to your organisation? Please describe all below.*

Please provide links where possible

3a. Which sector does this data policy relate to in general?*

- □ Internationally agreed policy
- □ Common nationally agreed policy
- □ Internal policy of an organisation
- □ Internal policy of a research institute
- □ Internal policy for a specific community
- □ Other, please specify _____

3b. Please, provide further details for the question above.

Which type of organisation/group/people does the policy apply to?

4a. What is the purpose of the data policy?*

- □ Data preparation
- □ Digital preservation
- \Box Data sharing
- □ Open access
- \Box Authorized access
- □ Other, please specify _____

4b. Please, provide further details for the question above.



- □ European Commission
- □ Ministry
- \Box International institute
- \Box National institute
- □ Strategic board
- □ Committee
- □ Dedicated task force
- □ Other, please specify _____

5b. Please, provide further details for the question above



6. What type of data is relevant to, or covered by the policy?*

- \Box Raw data
- \Box Published data
- \Box Data provided by this particular body
- \Box Not applicable
- □ Other, please specify _____

7. Who owns the data?*

- □ Depositor
- □ Repository
- □ Funder
- \Box Not applicable
- □ Other, please specify _____

8. For how long should the data be preserved, where timeframes are given/expected?*

- O Max. 5 years
- \bigcirc 5 20 years
- \bigcirc 20 50 years
- \bigcirc 50 100 years
- O Over 100 years
- \bigcirc Not decided
- O Not known
- \bigcirc Not applicable

9. Is there a timeframe for providing access to data? *

- \bigcirc The data is immediately accessible
- \bigcirc $% \left({{\rm{There}}} \right)$ is a moratorium of one year
- \bigcirc There is a moratorium of 2 years
- \odot $\,$ There is a moratorium of 2-5 years
- \bigcirc There is a moratorium of more than 5 years
- O Varying as dependent upon data
- \bigcirc The data will never be openly available
- \bigcirc Not applicable
- Not known

10. How is the policy maintained? When was the policy last updated?*

Also, describe any data management mechanisms in place.

11a. Are licenses recommended in the policy?*

Date: 2013-12-31D35.1Project: APARSENDoc. Identifier: APARSEN-REP-D35_1-01-1_0



- \Box Public domain
- □ Open licenses (e.g. Creative Commons)
- □ National licenses
- □ International licenses
- □ Uniform collection of licenses of various types
- □ Copyright protected, no licenses
- \Box Not applicable
- □ Other, please specify _____

11b. Please, provide further details for the question above.

Is ownership/usage/re-usage being discussed?

12. Are responsibilities for managing data policy defined within the policy? If yes, please specify:*

13a. Is there dedicated funding for the actions/activities described in the policy?*

- \Box Annual public funding
- \Box Funded by public financiers
- \Box Funded by commercial financiers
- \Box Funded by sales activities
- □ Short-term project-based funding
- \Box Not applicable
- □ Other, please specify _____

13b. Please, provide further details for the question above.

For example, are there any special administrative activities needed to get the funding?

14a. Who checks if involved parties follow the guidelines/policy?*

- \Box European Commission
- \Box Ministry
- \Box International institute
- \Box National institute
- \Box Strategic board
- \Box Not applicable
- □ Other, please specify _____

14b. Please, provide further details for the question above.

For example, what happens if the involved parties don't follow the policy?

Date: 2013-12-31 D35.1 Project: APARSEN Doc. Identifier: APARSEN-REP-D35_1-01-1_0



15. Is the policy linked to any research/career assessment? Any other incentives given?*

16. Are there any existing legal requirements for the policy, for example, on a national level? If so, please provide details.*

17. How does your data policy meet with different international or national standards, for example auditing methods or certifications?*

18a. Are there more specific requirements given in regards to the preservation activities?*

- \Box OAIS model
- □ File formats (e.g. ODF, TIFF, WAV, ...)
- \Box File categories (e.g. documents, audio, images, ...)
- \Box File size limits
- \Box Persistent identifiers
- \Box Not applicable
- □ Other, please specify _____

18b. Please, provide further details for the question above.

19a. Please specify any metadata requirements.*				
	Mandatory	Recommended	Optional	Not applicable
Common metadata format (e.g. METS, PREMIS)				
Descriptive metadata				



Provenance metadata		
Technical metadata		
Rights metadata		
Structural metadata		
Other metadata, please specify		

19b. Please, provide further details for the question above.

20. Please provide any other information which may be of relevance to current data policies?

GOVERNANCE STRUCTURE

In this section, we would like to know more about your governance structures. Proper governance structures and guidelines enable agreements for distributed heterogeneous data facilities, which in turn promotes data sharing.

21. Organisation profile *

Select from the following list the category which best describes your organisation?

- □ National library/archive
- □ University
- \Box Research institute
- \Box Research data archive
- □ Membership organisation
- □ Commercial supplier
- □ Government
- □ Funding organization
- □ Other, please specify ____

22. In which country does your organisation reside? *

23. How would you describe your role in relation to data? *

- □ Data provider
- \Box Data user
- □ Data curator
- \Box Not applicable
- □ Other, please specify _____



24. Please briefly explain your organisation's governance structure in relation to data policy management. *

- $\hfill\square$ Organisation has, e.g., a steering group responsible for the policy management
- □ Different parts of the policies are managed in different levels of the organisation
- $\hfill\square$ The policies are defined in each working group separately
- \Box Not applicable
- □ Other, please specify _____

25. Please provide any other information which may be of relevance to governance structure?

INTEROPERABILITY

Interoperability issues in data management are important to ensure that the preserved data are reachable and useful to others enabling cross-disciplinary research. We would like to know how your data policies and/or governance structures are, or can be, shared between different data repositories.

26. Do you think that any of your data policies are relevant to other organisations? If so, please provide details.*

27. Do you think that your governance structure is relevant to other organisations? If so, please provide details.*

28. Do you think that any of your data management mechanisms could be applied by other organisations? If so, please provide details.*

29. Do you have any of co-operation agreements in place with other institutions? If so, please provide details.*

Do you have any other feedback which you would like to provide to the APARSEN project?

