



A report on the range of policies required for and related to digital curation

Sarah Jones
Digital Curation Centre, University of Glasgow

DCC POLICIES REPORT
Deliverable H1.1

Version: 1.2
Date: 30 March 2009

Catalogue entry

Title	A report on the range of policies required for and related to digital curation
Creator	Digital Curation Centre, University of Glasgow
Subject	policies; curation; Research Council requirements; higher education; data management;
Description	This report investigates policy requirements related to digital curation, for example in the form of Research Council mandates, and current provision within higher education institutions to meet these. Recommendations are provided on key areas for policy development.
Publisher	Digital Curation Centre
Author	Sarah Jones
Date	2009.03.30
Type	Text
Format	Adobe Portable Document Format
Language	English
Rights	© Digital Curation Centre

Citation Guidelines

Jones, S., *A report on the range of policies required for and related to digital curation*, version 1.2, (DCC, Glasgow, March 2009)

Copyright Notices

Text and diagrams © Digital Curation Centre

Contents

1. EXECUTIVE SUMMARY.....	4
2. INTRODUCTION.....	5
2.1 BACKGROUND RESEARCH	5
3. REQUIREMENTS FOR DIGITAL CURATION	6
3.1 OVERVIEW	6
3.2 REQUIREMENTS BY UK RESEARCH FUNDER	8
<i>Table 1: Curation policies and support services of the main UK research funders</i>	<i>8</i>
3.2.1 AHRC – Arts and Humanities Research Council	11
3.2.2 BBSRC - Biotechnology and Biological Sciences Research Council	13
3.2.3 EPSRC - Engineering and Physical Sciences Research Council.....	15
3.2.4 ESRC - Economic and Social Research Council.....	16
3.2.5 MRC - Medical Research Council.....	18
3.2.6 NERC - Natural Environment Research Council	20
3.2.7 STFC - Science and Technology Facilities Council	21
3.2.8 The Wellcome Trust.....	23
4. POLICIES RELATED TO DIGITAL CURATION.....	25
4.1 OVERVIEW	25
<i>Figure 1: DCC Lifecycle Model mapped with policies and stakeholders.....</i>	<i>26</i>
4.2 CURRENT POLICY PROVISION	27
4.2.1 Repository & data centre policies	27
4.2.2 Data management and sharing plans	27
4.2.3 Institutional curation policies.....	28
4.3 ASPECTS TO CONSIDER WHEN CREATING CURATION POLICIES	29
4.4 SUMMARY AND RECOMMENDATIONS	30
5. CONCLUSION	31
6. BIBLIOGRAPHY	32

1. Executive Summary

A number of calls for open access to research outputs have been issued in recent years, such as those of the OECD and RCUK.¹ As open access is itself premised on effective curation, these developments have spurred policy creation. Research Councils recognise that curation begins with the initial research idea and most consequently now require a data management and sharing plan as part of applications. In addition the revised RCUK code for good research conduct is likely to encourage more institutional-level curation policies. This report maps the current curation policy landscape to identify areas requiring most development and support. The DCC will provide tools, such as policy templates, and guidance documents to assist those creating new policies.

All of the research funders assessed in this report have a policy on access to publications in line with the joint RCUK statement. Most also have a data policy. The level of detail in these varies significantly, however most funding body policies cover similar ground as demonstrated in *Table 1*. The main difference is the level of guidance and support infrastructure provided – many funders expect researchers to draw on institutional provision, which can be lacking.

Provision of curation policies within HE institutions is patchy. Several reports have recently issued concerns that few institutions have policies in place to guide the creation, maintenance and long-term preservation of digital resources.² Data management and sharing plans are becoming more commonplace due to funders' mandates, however there is still room for development and a clear need for support. Repository and data centre policies are the most advanced and tools, such as the OpenDOAR policy tool, and guidance from organisations such as the Repositories Support Project has been provided for some time.³

Several recommendations are made to assist researchers and institutions engaged in creating new curation policies. These focus on utilising existing tools and guidance, ensuring associated policies work in harmony, exploiting existing structures and workflows, and assisting uptake through testing, incentivisation and provision of support. Three broader issues to be addressed collaboratively by the entire sector are also noted, namely: identifying roles and responsibilities; assessing costs and benefits to determine how curation should be financed; and developing a robust and sustainable curation infrastructure. Addressing these will provide the necessary underpinning to implement curation policies.

1 The OECD principles are at: <http://www.oecd.org/dataoecd/9/61/38500813.pdf> The RCUK 2005 draft position is at: <http://www.rcuk.ac.uk/research/outputs/access/2005.htm> and the 2006 update: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf>

2 See for example Beagrie, Neil et al, *Digital preservation policies study*, (2008), available at: http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

3 See: <http://www.opendoar.org/tools/en/policies.php> and <http://www.rsp.ac.uk/repos/rules>

2. Introduction

This report was developed in response to feedback from training events and the DCC Associates Network. A central dissemination point with advice and guidance on the curation policies of UK funding bodies was requested. The DCC has drawn this information together and provided guidance on the next steps needed to enhance the existing curation policy landscape. Developing and implementing a range of curation policies is essential to effectively manage, discover and reuse research outputs. To encourage this, the DCC is creating generic policy templates that can be downloaded and customised to meet specific user needs. This preliminary study helps to define areas of need. The report sets out to:

1. provide an overview of curation requirements as set out by UK research funders;
2. examine the current curation policy landscape to identify gaps in provision;
3. recommend policy developments to ensure appropriate curation of research outputs.

As digital curation encompasses a variety of activities across the information lifecycle, a range of policies is required. These will inevitably overlap and address the concerns of various stakeholders involved in creating and managing research outputs. Section 3 of this study investigates the policies required by assessing the curation and access mandates of the main UK research funders. The requirements are then contrasted with existing curation policies within higher education in Section 4 to identify gaps and areas requiring development.

The policy landscape is rapidly changing so details in this report can only provide a snapshot of the current position. The information provided here has been abstracted for DCC curation policy web pages, which act as an overview to each funders' policy and a portal for related events, tools and guidance.⁴ The information provided online will be updated regularly, so should be consulted over this report in the longer term. This report notes likely future developments, such as Research Council plans to harmonise their data policies by late 2009, which will be monitored and communicated through the policy web pages.

2.1 Background research

Several previous studies, which comment on issues of curation policy, have been built on in this report. In the JISC-commissioned *Dealing with Data* report, Liz Lyon provides an overview of the policy drivers for open access and contrasts each Research Council's stance on maintaining publications and data.⁵ The Research Information Network produced a detailed report on research funders' policies with respect to the management of information.⁶ This considers mandates on published outputs, theses, data and grey literature, setting requirements in the wider context of repository provision. A more recent study of relevance was the *Digital Preservation Policies Study* commissioned by the JISC in 2008.⁷ This study responds to the lack of institutional level curation policies across the HE sector by providing a framework for those planning to create such policies, together with a series of mappings to core university business drivers so preservation policies can be embedded in wider aims. A DCC / JISC comparative study into international approaches to data sharing⁸ has also been considered. This builds on the DataShare *State of the Art Review* to consider what policies

4 See: <http://www.dcc.ac.uk/resource/curation-policies/>

5 Lyon, Liz, *Dealing with Data: Roles, Rights, Responsibilities and Relationships*, (2007), available at: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

6 RIN, *Research Funders' Policies for the management of information outputs*, (2007), available at: <http://www.rin.ac.uk/files/Funders%27%20Policy%20&%20Practice%20-%20Final%20Report.pdf>

7 Beagrie, Neil et al, *Digital preservation policies study*, (2008), available at:

http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

8 Ruusalepp, Raivo, *A comparative study of international approaches to enabling the sharing of research data*, (2008), available at:

http://www.dcc.ac.uk/docs/publications/reports/Data_Sharing_Report.pdf

are in place for data sharing across OECD countries, as well as assessing the infrastructure and support currently provided by data centres, repositories and domain specific initiatives.

Various services and initiatives have also been taken into consideration here, such as the OpenDOAR policy tool and SHERPA JULIET and RoMEO services. Guidance from UK and international organisations such as the Repository Support Programme, UK Data Archive, Australian National University and MIT, has also been consulted. References to these sources is provided in the bibliography and through the DCC curation web pages.

3. Requirements for digital curation

3.1 Overview

A number of agenda-setting reports and statements have been released in response to the growing levels of digital research outputs. These call for a co-ordinated approach to curation, improved infrastructure, secure investment, and open access. In response UK funding bodies have been issuing policies related to digital curation and access, which make requirements for researchers and their institutions explicit. Central to these requirements is the need for a more robust approach to creating, maintaining and preserving the outputs of research to ensure they can be shared and reused. Indeed, one of the RCUK expectations for societal and economic impact is that those who receive funding “take responsibility for the curation, management and exploitation of data for future use.”⁹ Maintaining data is also crucial to allow for research validation, which is a commonly accepted principle of good scientific practice.¹⁰ As such, curation should be central to the work of HE institutions.

The open access movement, which has been gaining pace since the 2001 Budapest Open Access Initiative, advances the curation agenda as sustained access to research outputs is itself premised on effective creation and management practices. Cases in which research outputs cannot be made available are often due to failings in the original capture or storage procedures, for example inadequate rights agreements, lack of documentation or data loss. Moves to increase access to research outputs naturally raise curation questions. Significant developments in the open access field of late have been the joint RCUK statement on open access to research outputs¹¹ and the OECD principles for access to research data.¹² Community policies such as the Bermuda Statement, which advocates genomic sequence information be released immediately and freely in the public domain, add further weight to calls for open access and provide useful models of how curation can become the norm.¹³

These developments have led UK research funders to release policies on curation and data sharing. The RCUK statement on open access to research outputs proposes that a copy of any published journal articles or conference proceedings resulting from Research Council funded research should be deposited in an appropriate e-print repository at the earliest opportunity. This move to promote wider access underpins the common belief that publicly funded research must be made available to the public.¹⁴ Each Research Council has

9 RCUK, *Expectations for Societal and Economic Impact*, available at:

<http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/innovation/expectationssei.pdf>

10 See for example: European Science Foundation, *Good Scientific Practice in Research and Scholarship*, policy briefing 10 (2000), at: <http://www.esf.org/publications/policy-briefings.html>

11 The 2005 draft position statement is at: <http://www.rcuk.ac.uk/research/outputs/access/2005.htm> and the 2006 update at: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf>

12 OECD, *Principles and Guidelines for Access to Research Data from Public Funding*, (2007), available at: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

13 Bentley, D. R. (1996). “Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain” *Science Magazine*, 274, 533-534, available at: <http://www.sciencemag.org/cgi/content/full/274/5287/533>

14 See the core principles (section 3) in the RCUK position statement and the OECD central principle as noted in OECD, *Promoting Access to Public Research Data for Scientific, Economic, and Social Development*, (2003), available at: http://dataaccess.ucsd.edu/Final_Report_2003.pdf

subsequently released its own position statement based on this joint agreement. The OECD principles for access to research meanwhile stem from a declaration first agreed in 2004.¹⁵ These have since been extended and guidance added. The OECD principles are: openness; flexibility; transparency; legal conformity; protection of intellectual property; formal responsibility; professionalism; interoperability; quality; security; efficiency; accountability; and sustainability. The additional guidelines note that three areas still require further attention: roles and responsibilities; determining how curation should be funded; and appreciating benefits to be gained so we can ensure cost-effectiveness.

A synthesis of the RCUK statement and OECD principles has been provided by the RIN to avoid inconsistencies and duplication.¹⁶ Five principles required for effective stewardship of digital research data are asserted:

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

In order to achieve the goals of the open access movement and comply with curation policies, a supporting infrastructure is required. The OECD guidelines recognise international frameworks to facilitate access are still lacking in many countries and suggest research institutions and government organisations take responsibility for long-term sustainability. If the policies currently being developed are to be effectively implemented, infrastructure issues should be addressed simultaneously.

Table 1 overleaf summarises the coverage of main UK research funders' policies and related support infrastructure. A brief cross-council overview of each policy area is provided before the more detailed description of each funding body's policy. The SHERPA JULIET service, which provides details of the open access archiving and publication policies of various funding bodies worldwide, may also be of interest.¹⁷

15 See Annex 1: Declaration on access to research data from public funding, available at:

http://www.oecd.org/document/15/0,3343,en_2649_34487_25998799_1_1_1_1,00.html

16 RIN, *Stewardship of digital research data: a framework of principles and guidelines*, (2008), available at:

<http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20full%20version%20-%20final.pdf>

17 See: <http://www.sherpa.ac.uk/juliet/>

3.2 Requirements by UK research funder

Research Funders	Policy coverage		Curation stipulations					Support provided		
	Published outputs	Data	Time limits	Data plan	Access / sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre
AHRC - Arts and Humanities Research Council	●	●	●	●	●	●	○	●	○	◐
BBSRC - Biotechnology & Biological Sciences Research Council	●	●	●	●	●	●	●	◐	●	◐
EPSRC - Engineering and Physical Sciences Research Council	◐	○	◐	○	◐	○	○	○	○	◐
ESRC - Economic and Social Research Council	●	●	●	●	●	●	●	●	●	●
MRC - Medical Research Council	●	●	●	●	●	●	○	●	●	◐
NERC - Natural Environment Research Council	●	●	●	●	●	●	●	●	●	●
STFC - Science and Technology Facilities Council	●	○	●	○	◐	○	○	○	●	◐
Wellcome Trust	●	●	●	●	●	●	◐	○	●	◐

Table 1: Curation policies and support services of the main UK research funders

Key: ● full coverage ◐ partial coverage ○ no coverage

Table 1 terminology clarifications

Published outputs: a policy on published outputs e.g. journal articles and conference papers

Data: a data policy or statement on access to and maintenance of electronic resources

Data plan: requirement to consider data creation, management or sharing in the application

Time limits: set timeframes for making content accessible or preserving research outputs

Access / sharing: promotion of OA journals, deposit in repositories, data sharing or reuse

Monitoring: whether compliance is monitored or action taken such as withholding funds

Curation: stipulations on long-term maintenance and preservation of research outputs

Guidance: best practice guides or curation support staff available to funded researchers

Repository: provision of a repository to make published research outputs accessible

Data centre: provision of a data centre to curate unpublished electronic resources or data

Published outputs

The research funders' policies on published outputs are aligned with the joint RCUK position statement, which was first issued in June 2005.¹⁸ All advocate open access to outputs from their funded research programmes and many provide a repository service in support of this requirement. There are differences regarding how publications fees should be met, and some funders include additional stipulations – NERC, for example, may take compliance with this policy into account when considering further applications for funding.

Data

Most funders have some form of policy regarding data, however the extent and coverage of these vary greatly. In several cases researchers are directed to good practice guides, which provide recommendations on documenting and maintaining research. There are only two Research Councils without a formal data policy as yet – the EPSRC and STFC – though a data policy is currently being developed at EPSRC and STFC appears to continue CCLRC and PPARC procedures.

Time limits

The timeframes stipulated for access and curation vary by funding body. Most expect publications to be made openly available as soon as possible or in a timely manner, which is generally understood to be at least within six months of publication of results. The ESRC and AHDS (only in the case of archaeology) expect deposit of data within three months of the end of the award, and also advocate a relationship with the data centre from the outset of the project. The BBSRC, MRC and Wellcome Trust have a general statement that data must be kept securely for a period of ten years after the completion of a research project, while the EPSRC maintains it should be held for an 'appropriate' length of time.

Data plan

Most research funders require applicants to submit a statement on access, management and long-term curation of their research outputs at the proposal stage. The focus of this statement varies by funder: the AHRC, ESRC and NERC all require a statement on how resources will be created so they can be preserved in the long-term, while the BBSRC, MRC and Wellcome Trust focus heavily on the data sharing potential of research resources. Neither the EPSRC nor STFC currently require applicants to submit data sharing or curation plans as part of the proposal, however this is likely to change as data policies emerge.

Access / data sharing

All funders have signed up to the RCUK statement on open access of research outputs and advocate making publications widely accessible. They largely agree to meet publication fees, normally as indirect costs, to ensure research is freely accessible. The MRC and Wellcome Trust also encourage - or in cases where they have paid publication fees, require - licences that allow articles to be freely copied and reused for purposes such as text and data mining. Some moves are also being made towards linking publications with source data;¹⁹ for example UK PubMed Central allow deposit of supplemental material.

The concept of open data is not advocated in any of the research funders' policies, however the BBSRC, MRC and Wellcome Trust have the strongest ethos of data sharing, expecting

18 See RCUK website at: <http://www.rcuk.ac.uk/access/default.htm>

19 The JISC funded StORe (Source to Output Repositories) project covered this, see: <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/store.aspx>

data to be made available with as few restrictions as possible. The ESRC and NERC facilitate data sharing through their funded data centres, however licence fees and access restrictions are often applied as their remit is to serve research and teaching communities. The AHRC provides access to, and a cross-search of, their funded archaeology data through the ADS and requires other award holders to keep data accessible for a minimum of three years. The STFC has not provided a clear statement on expectations for data sharing. According to a RIN study the EPSRC considers the discipline areas it covers do not have so much need for data sharing,²⁰ however the policy that is currently being developed may revise this position.

Long-term curation

Most of the funders consider curation in detail in their policies. The AHRC, BBSRC, ESRC, MRC, NERC and Wellcome Trust all consider various aspects of the curation lifecycle, for example noting the need to create resources according to appropriate standards and best practice, maintain adequate documentation and metadata to ensure usability, and manage data appropriately in the short-term so it can be preserved for the future. The EPSRC only has one stipulation – that data be appropriately stored for a minimum of 10 years – while the STFC does not appear to have any formal requirements addressing curation.

Monitoring

NERC and the Wellcome Trust note they monitor compliance with the open access policy on publications. The BBSRC will monitor adherence to the data management and sharing plan and may take this into consideration for future proposals, while the ESRC could withhold the final grant payment if data is not deposited on time. The extent to which such penalties are applied is unclear. The other funders meanwhile do not appear to monitor adherence or impose penalties for non-compliance with their curation policies.

Guidance

The extent to which guidance and support services are provided varies significantly. The best served researchers are those funded by the ESRC and NERC. Support staff based at the data centres will provide assistance and advice on best practice throughout the award. The AHRC runs a similar service for archaeology researchers and has legacy guides online for researchers in other fields. The MRC meanwhile is setting up a data support service and already provides some best practice guides and data sharing toolkits. BBSRC does not appear to have much guidance online, but states that information on relevant standards and best practice will be provided and a main contact for this is listed. No particular sources of guidance were noted by the EPSRC, STFC and Wellcome Trust in their policies or found on their websites. It may be that curation support is offered less formally by these funding bodies.

Repository

Most research funders provide a publications repository for their funded researchers. ESRC, NERC and STFC all run their own e-Prints service while BBSRC, the MRC and Wellcome Trust are partners in PubMed Central. The only Councils that do not provide a repository for published outputs are the AHRC and EPSRC. Researchers supported by these Councils are expected to use any institutional or subject based repositories available to them.

Data Centre

Provision of data centres is patchy - very few funding bodies have a full data service in place. The exceptions are the ESRC and NERC, which both provide comprehensive preservation and support services. The BBSRC, MRC and Wellcome Trust meanwhile agree the cost of long-term curation can be included in the original proposal. The AHRC provides a data service for researchers in the area of archaeology through ADS and it appears STFC have several services and agreements in place to provide pockets of support, for example through the UK Solar System Data Centre. The Wellcome Trust, BBSRC, MRC and EPSRC all contribute to the European Bioinformatics Institute,²¹ however for research that falls outside the EBI remit the institutions in which funded researchers are based are expected to maintain outputs in the long-term.

20 RIN, *Research funders' policies for the management of information outputs*, (2007), p61, available at: <http://www.rin.ac.uk/files/Funders%27%20Policy%20&%20Practice%20-%20Final%20Report.pdf>

21 For more information on EBI, see: <http://www.ebi.ac.uk/>

3.2.1 AHRC – Arts and Humanities Research Council

The AHRC has released a statement on open access to research outputs in line with the RCUK position.²² The Council's policy on electronic resources comprises a few clauses in the *Research Funding Guide*, last published November 2008.²³

Policy stipulations

Overview

- The AHRC provides expected timeframes for data release and availability: electronic resources must remain accessible for a minimum of three years after the end of the award and deposits to the Archaeology Data Service (ADS)²⁴ must be made within three months of project completion.
- If applicants plan to create significant electronic resources, they are expected to complete a technical appendix at the proposal stage. This has six sections: project management of technical aspects; data development methods; infrastructural support; data preservation and sustainability; access; and copyright and IPR issues.
- The AHRC requires all research outputs to be made available in an accessible repository. Electronic resources need to remain available for at least three years.
- The details required by the technical appendix cover data creation and sustainability, however there is not a specific mandate to preserve apart from in the case of archaeology, where grant holders are expected to deposit resources with the ADS.
- Compliance with the curation policy does not appear to be actively monitored, nor are penalties stated for failure to fulfil these requirements.

Published outputs

The AHRC does not prescribe where authors publish their research, nor note a preference for open access journals. If a journal is chosen which imposes page charges or other publishing fees, it is for authors' institutions to decide whether they are prepared to meet these. Such funds could be part of an institution's indirect costs under the full economic costing regime.

The AHRC requires that funded researchers:

- ensure deposit of a copy of any resultant articles published in journals or conference proceedings in appropriate repository
- wherever possible, ensure deposit of the bibliographical metadata relating to such articles, including a link to the publisher's website, at or around the time of publication

Full implementation of these requirements must be undertaken such that current copyright and licensing policies, for example, embargo periods and provisions limiting the use of deposited content to non-commercial purposes, are respected by authors.

22 <http://www.ahrc.ac.uk/FundingOpportunities/Documents/access%20to%20research%20outputs.pdf>

23 Arts and Humanities Research Council, *Research Funding Guide*, (2008), particularly sections on deposit of resources or datasets and self-archiving policy on p86, available at:

<http://www.ahrc.ac.uk/FundingOpportunities/Documents/Research%20Funding%20Guide.pdf>

24 See <http://ads.ahds.ac.uk/>

Data

If a project results in the creation of an electronic resource, grant holders must:

- complete a technical appendix as part of the application to allow the technical feasibility of the project to be assessed by the Peer Review College;
- make any significant electronic resources or datasets available in an accessible depository for at least three years after the end of their grant.

The AHRC provides a repository and support service for archaeology researchers through the ADS. Grant holders in this field must:

- consult with the ADS within three months of the start of the project to agree the form and extent of electronic materials to be deposited;
- offer significant electronic resources or datasets, together with documentation, to the ADS within three months of the end of the project;
- agree a waiver of deposit with the ADS in cases where the offer of deposit is not accepted or a specified reason prevents deposit.

When a research project results in a non-permanent output such as an installation or performance, award holders are required to send one copy of any publication or public output (e.g. CD-ROM, video, photographs etc) to the AHRC.²⁵

Support provided

- Support and guidance for archaeology researchers is available through ADS staff, who provide assistance from the proposal stage, through creating sustainable resources to depositing them for long-term preservation. Legacy guidance materials such as subject specific preservation handbooks and information papers are still available through the AHDS web pages for other researchers.²⁶ It should be noted that these are not being updated so will lose currency.
- A publications repository is not provided by the AHRC. Researchers are expected to make use of institutional and subject-based repositories available to them.
- The AHRC provides a comprehensive data support service for archaeology researchers through the ADS. Such support for other researchers ceased with the end of funding for the Arts and Humanities Data Service (AHDS) in Spring 2008. The AHRC expects non-archaeology researchers to draw on other data centres and repositories to ensure continued access to research outputs for at least three years after the end of the award.

25 This stipulation applies to three schemes only: Fellowships in the Creative and Performing Arts; Research Grants – practice-led and applied route; and Research Leave.

26 For example see the Guides to Good Practice, Information Papers and Case Studies available at: <http://www.ahds.ac.uk/about/publications/index.htm> or the Preservation Handbooks based on data types at: <http://ahds.ac.uk/preservation/ahds-preservation-documents.htm>

3.2.2 BBSRC - Biotechnology and Biological Sciences Research Council

BBSRC has a statement on access to published research outputs,²⁷ as well as an extensive *Data Sharing Policy*, which came into effect in April 2007.²⁸ The *Statement on Safeguarding Good Scientific Practice* acts as an overarching framework for these policies.²⁹

Policy stipulations

Overview

- BBSRC policies stipulate time limits for sharing research outputs: publications must be deposited at the earliest opportunity and data must be made available in a timely and responsible manner. It is expected that timely release would generally be no later than the release through publication of the main findings, or three years as a general guide. Data should be maintained for 10 years after project completion.
- Researchers are required to submit a data sharing plan as part of the proposal. This may include details of: data areas, types and formats; standards and metadata; secondary use; methods for data sharing; and timeframes for release.³⁰
- The Council actively promotes data sharing and encourages researchers to make material openly accessible. A publications repository and financial support for data sharing is available to facilitate sustained access.
- The coverage of the *Data Sharing Policy* and the *Statement on Safeguarding Good Scientific Practice* is testament to BBSRC's commitment to the entire lifecycle of research outputs. Researchers are expected to keep data securely for ten years after the end of a project through their institutions.
- Adherence to the proposed data management and sharing plan will be monitored through the final report assessment procedure and may be taken into account when assessing future proposals.

Published outputs

BBSRC's statement on published outputs is in line with the agreed RCUK stance. Where journals charge authors a fee for publishing open access articles, this can be met as part of the grant. Grant holders are expected to:

- deposit a copy of any resulting published journal article or conference proceedings, at the earliest opportunity, in PubMed Central or an appropriate e-print repository;
- wherever possible, deposit the bibliographical metadata (including a link to the publisher's website), at or around the time of publication, in the relevant repository.

Full implementation of these requirements requires that current copyright and licensing policies, such as embargo periods, are maintained by publishers and respected by authors.

27 BBSRC, *Policy on access to research outputs*, (2008), available at: http://www.bbsrc.ac.uk/publications/policy/access_research_outputs.html, also see details in *BBSRC Research Grants: the guide*, Version 8.04, (2008), clause AC1, p45, available at: http://www.bbsrc.ac.uk/funding/apply/grants_guide.pdf

28 BBSRC, *BBSRC's Data Sharing Policy*, (2007), available at: http://www.bbsrc.ac.uk/publications/policy/data_sharing_policy.pdf

29 BBSRC, *BBSRC Statement on Safeguarding Good Scientific Practice*, available at: http://www.bbsrc.ac.uk/publications/policy/good_scientific_practice.pdf

30 See 'Data Sharing Areas' on p6 of the *BBSRC Data Sharing Policy*

Data

Researchers are encouraged to follow best practice through the curation lifecycle to ensure appropriate creation, management and reuse of data. The main requirements from the *Data Sharing Policy* and *Statement on Safeguarding Good Scientific Practice* are as follows:

- Applicants must submit a statement on data sharing to be assessed by reviewers. This should include concise plans for data management and sharing or provide explicit reasons why data sharing is not possible or appropriate.
- Researchers should make use of current best practice and generate data and documentation using widely accepted formats, methodologies and standards.
- Data should be accompanied by contextual information (documentation / metadata) 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.
- Data should be made available through existing community resources or databases where possible.
- BBSRC expects research data to be made available for subsequent research with as few restrictions as possible in a timely and responsible manner. Timely release would generally be no later than publication of the main findings and should be in-line with established best practice or within three years if no best practice exists.
- Data must be kept securely in paper or electronic form for a period of ten years after the completion of a research project.
- Researchers are expected to ensure appropriate data management strategies are in place throughout the research project.
- Institutions receiving BBSRC funding must have guidelines setting out responsibilities and procedures for keeping data.

Support provided

- A main point of contact for data sharing queries is provided³¹ and the Council states it will provide guidance on existing standards, guidelines, databases and resources that may be relevant.
- Repository support for publications and accompanying material is available through UK PubMed Central.³²
- Central repository support for data is not provided, however BBSRC recognises data sharing has time and cost implications, so funding for this can be requested as part of the full economic cost of a research project. The Council also supports the European Bioinformatics Institute.

31 See David McAllister's contact details on p18 of the BBSRC *Data Sharing Policy*

32 See: <http://ukpmc.ac.uk/>

3.2.3 EPSRC - Engineering and Physical Sciences Research Council

EPSRC has recently mandated open-access publication of research that it funds.³³ Further details will be provided shortly, but for the moment the Council links through to the joint RCUK statement.³⁴ EPSRC does not have a data policy at present, however a few stipulations on maintaining data are made in the *Guide to Good Practice in Science and Engineering Research*.³⁵ The Council is currently developing policies on maintaining research outputs.

Policy stipulations

Overview

- The only time constraint stipulated by EPSRC at present is that data should be securely stored for an appropriate time. The Council's agreement with the RCUK statement also suggests that publications should be deposited in an accessible repository at the earliest opportunity, preferably at or around the time of publication.
- EPSRC does not require researchers to provide a statement on data management or sharing as part of the application process.
- On its website EPSRC states it is 'strongly committed to the principles outlined in the RCUK position statement on access to research outputs.'³⁶ According to research conducted by the RIN, the Council considers that the discipline areas it covers do not have so much need for data sharing, and that in most cases a journal publication summarising results is sufficient without providing access to the original dataset.³⁷ The position with respect to data sharing may be revised in the planned data policy.
- Few stipulations are made regarding long-term curation of research outputs, aside from a principle of good scientific practice being to securely store primary data. The fact that much of the data the Council funds could potentially be recreated in the future may underlie this current position.
- Adherence with EPSRC's curation requirements does not appear to be monitored.

Published outputs

EPSRC agreed to mandate open access publication in December 2008. Academics should be able to choose whether they use the so-called green option (i.e. self-archiving in an on-line repository) or to use the gold option (i.e. pay-to-publish in an open access journal). Full policy details will be published in spring 2009.

Data

One of four EPSRC principles of good scientific practice is securing and storing primary data:

Primary data as the basis for publications should be securely stored for an appropriate time in a durable form under the control of the institution of their origin.

33 See the news story on this change from *Research information*, 12th February 2009 at:

http://www.researchinformation.info/news/news_story.php?news_id=436

34 For more information see: <http://www.epsrc.ac.uk/AboutEPSRC/AccessInfo/ROAccess.htm>

35 EPSRC, *Guide to Good Practice in Science and Engineering Research*, (2006), available at: <http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/GoodPracticeGuideSciEngRes.pdf>

36 See: <http://www.epsrc.ac.uk/AboutEPSRC/AccessInfo/ROAccess.htm>

37 RIN, *Research funders' policies for the management of information outputs*, (2007), p61, available at: <http://www.rin.ac.uk/files/Funders%27%20Policy%20&%20Practice%20-%20Final%20Report.pdf>

EPSRC expects that institutions will have policies in place to ensure sound codes of practice are respected for work they undertake, and that even if the individual researchers responsible for generating data relocate, a set will be maintained in the institution of origin.

Support provided

- As the EPSRC's policy regarding information outputs is minimal, little guidance or support appears to be offered by the Council.
- No repository support is provided for publications or data. It is expected researchers will utilise the institutional or subject-based repositories available to them.
- The responsibility to securely store data falls to the institution awarded the grant, however the EPSRC does provide support to the European Bioinformatics Institute so bioinformatics data could be added to community resources maintained there.

3.2.4 ESRC - Economic and Social Research Council

The ESRC has released a statement on open access of published research³⁸ and the current datasets policy is incorporated as an appendix in the *Research Funding Guide*.³⁹ The original, more extensive *Data Policy*⁴⁰ is currently under review, and will apparently be widened in scope under the new title *Research Resources Policy*.⁴¹

Policy stipulations

Overview

- Time limits are stipulated for releasing data and publications: grant holders are expected to deposit publications in the ESRC awards and outputs repository at the earliest opportunity and data must be offered to the Economic and Social Data Service based at the UKDA within three months of the end of the award.
- ESRC require applicants to consider what outputs will be created at the proposal stage and how these will be made available in the long-term. Five questions covering a survey of existing data, plans for archiving and potential users should be covered.⁴²
- Researchers are expected to make all outputs accessible as soon as possible. ESRC provides a publications repository and data service to facilitate this.
- The Council's commitment to long-term curation is displayed through the repository and data archive services it supports. Researchers are expected to seek guidance and follow best practice to ensure outputs are created and maintained appropriately.
- ESRC states it will monitor compliance with its policies. The final payment of a grant will be withheld if data has not been deposited to the required standard, unless a waiver of deposit has been agreed in advance.

38 See the ESRC statement at: <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Support/access/>

39 ESRC, *Research Funding Guide*, (2008), available at:

http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/ESRC%20Research%20Funding%20Guide_tcm6-9734.pdf

40 ESRC, *Data Policy*, (2000), available at:

http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/DataPolicy2000_tcm6-12051.pdf

41 As noted in Liz Lyon, *Dealing with Data*, (2007), p18

42 For details of what should be included see: <http://www.esds.ac.uk/aandp/create/esrcfaq.asp>

Published outputs

The ESRC is committed to the principles articulated in the RCUK statement on access to research outputs. Authors choose where to place their research, so ESRC believes it is for authors' institutions to decide whether they are prepared to use funds from full economic costs to pay any page charges or other publishing fees.

It is mandatory for grant holders, at the earliest opportunity, to:

- personally deposit, or otherwise ensure deposit, of any resultant articles published in journals or conference proceedings, in the ESRC awards and outputs repository;
- wherever possible, personally deposit, or otherwise ensure deposit of bibliographical metadata relating to such articles, including a link to the publisher's website, at or around the time of publication, in the ESRC awards and outputs repository.

Grant holders are also encouraged to submit copies of resultant publications and/or associated metadata with institutional and other appropriate repositories. Full implementation of these requirements requires that current copyright and licensing policies, such as embargo periods or provisions limiting the use of deposited content to non-commercial purposes, are respected by authors.

Data

A principle of good scientific practice for the ESRC is that primary data from research projects be properly secured and stored, including data deposit. As such it has a robust data policy and support infrastructure to help researchers comply with requirements. Grant holders must:

- offer for deposit copies of both machine-readable and non-machine-readable qualitative data to the ESDS within three months of the end of the grant;
- provide adequate accompanying documentation to a standard which would enable the data to be used by a third party;
- or in cases where deposit is not possible, agree a waiver of deposit with the ESDS team in advance of the end of the project.

It is recommended that applicants likely to produce a dataset contact the ESDS Acquisitions team at the UKDA prior to making their application to ensure adequate provision for preparation of data can be made in their application.

Support provided

- The ESRC offer extensive support to researchers creating digital records. Data centre staff will assist from the application process, throughout the project to final deposit and reuse. Various guides are available through the website and a helpdesk is run.⁴³
- The ESRC awards and outputs repository provides a place for grant holders to deposit their publications.⁴⁴
- Time can be set aside within the grant and funding requested for the preparation of data for deposit. Data will be preserved in the long-term by the UKDA.⁴⁵

43 See guidance at: <http://www.esds.ac.uk/aandp/create/createintro.asp>

44 See: <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/AdvancedSearchPage3.aspx>

45 Guidance for ESRC researchers is at: <http://www.esds.ac.uk/aandp/create/storefaq.asp>

3.2.5 MRC - Medical Research Council

The MRC has released a position statement in support of open and unrestricted access to published research in line with the RCUK position.⁴⁶ It also has a brief *Policy on Data Sharing and Preservation*.⁴⁷ More detailed guidance on how to plan, conduct and record research is available through the Council's *Good Research Practice Guide*.⁴⁸

Policy stipulations

Overview

- The MRC expects publications to be made openly accessible as soon as possible and in any event within six months of the journal publisher's official date of final publication. Data should be made available in a timely and responsible manner, and securely maintained for a minimum of ten years after completion of the research. A limited, defined period of exclusive use of data for primary research is reasonable.
- Applicants are expected to consider the future of their resources at the proposal stage through completion of a data sharing and preservation strategy. This should include a summary of the type of data to be generated, foreseeable research uses, and plans for preparing and documenting data for preservation for sharing. Applicants requesting funds to extend existing data sets should also explain how this adds value and how sharing would provide opportunities for coordination or collaboration.
- The MRC is a clear proponent of the open access and data sharing movement. It expects publications to be made available promptly and encourages researchers to make data available with as few restrictions as possible.
- Stipulations on long-term curation and preservation of research outputs are made in the *Policy on Data Sharing and Preservation* and *Good Research Practice Guide*. Recommendations are provided on gathering, recording and storing the data.
- It does not appear compliance with MRC's open access or data sharing policies is monitored as yet.

Published outputs

The MRC is a champion of open access publishing in science. It will support any necessary charges levied by publishers who offer open access options. Authors are expected to maximise the opportunities to make their results available for free and, where possible, to retain their copyright. The MRC requires that:

- electronic copies of any research papers accepted for publication in a peer-reviewed journal, supported in whole or in part by MRC funding, be deposited into PubMed Central (PMC) or UK PMC to be made freely available as soon as possible and in any event within six months of the journal publisher's official date of final publication;
- when an open access fee has been paid, authors and publishers licence research papers such that they may be freely copied and reused for purposes such as text and data mining, provided that such uses are fully attributed;
- where possible, published results should provide links to the associated data.

46 See MRC statement at: <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Openaccesspublishing/Positionstatement/MRC004368>

47 See the Data Sharing policy at: <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm>

48 MRC, *Good Research Practice Guide*, (2000), available at: <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002415>

Data

The MRC *Policy on Data Sharing and Preservation* requires researchers make clear provision for curating data when planning and executing their research. The *Good Research Practice Guide* specifies in more detail what is expected. The main requirements are that:

- all funding proposals must include a strategy for data preservation and sharing in the case for support;
- valuable data arising from MRC-funded research should be made available to the scientific community with as few restrictions as possible, in a timely and responsible manner;
- for medical research involving personal data, the appropriate regulatory permissions – ethical, legal and institutional – must be in place before the data can be shared;
- individuals or institutes that received MRC funding to create or collect the data must ensure it is properly curated throughout its lifecycle and released with the appropriate high-quality metadata;
- data should be backed-up regularly with duplicate copies held on disc in a secure but readily accessible archive;
- copies of relevant software, particularly the version used to process electronic data, must be retained along with the raw data to ensure future access;
- special attention should be paid to guaranteeing the security of electronic data.

Support provided

- The MRC provides support to researchers through data sharing toolkits and best practice guides.⁴⁹ A data support service is also being developed at present.
- Researchers are provided with the UK PubMed Central repository as a means to provide centralised access to their published research.
- The MRC encourages curation and long-term management of its research outputs and provides funds for preservation. The onus is on the Principal Investigator and his/her institution as a supporting preservation infrastructure is not yet offered, though a data support service is in development. The MRC also provides support for EBI.

⁴⁹ See MRC, *Principles for access to, and use of, MRC funded research data*, (2007) available at: <http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc003759.pdf> & activity update www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Recentactivities/index.htm

3.2.6 NERC - Natural Environment Research Council

NERC has released a position statement on access to published research.⁵⁰ As with the other Research Councils, this is in line with the RCUK statement. NERC also has an extensive *Data Policy Handbook*, which is currently under review.⁵¹

Policy stipulations

Overview

- Publications resulting from NERC funded research must be deposited at the earliest opportunity and data must be offered after a 'reasonable period' of exclusive use.
- Researchers are expected to consider aspects of data creation and management prior to beginning research then produce a more detailed data plan if funded. Overarching data plans will be produced for each thematic programme by NERC data centre staff. A data plan should describe the dataset and related documentation to be created, establish intellectual ownership, and determine roles and responsibilities for how the data will be created, stored, backed up, archived, accessed and reused.
- Publications and data must be made accessible through repositories and NERC's data centres. Minimum quality standards are enforced to ensure data can be reused and understood by third parties.
- Long-term curation is central to NERC and an extensive support infrastructure is in place to facilitate this.
- Compliance with the policy on publications may be considered in future applications. Note is not made of whether compliance with the data policy will be checked.

Published outputs

NERC is committed to the principles articulated in the RCUK statement on access to research outputs. As such, it requires that:

- an electronic copy of any published peer-reviewed paper, supported in whole or in part by NERC-funding, is deposited at the earliest opportunity in an e-print repository.

Full implementation of these requirements requires that current copyright and licensing policies, such as embargo periods, are maintained by publishers and respected by authors.

An e-print repository has been established to assist researchers to provide access to their publications and compliance with this policy will be taken into consideration when considering further applications for funding.

Data

NERC has had a data policy in place since the mid-1990s. The current version, dated 2002, is under review. The policy covers the entire data lifecycle from planning, through creation, preservation and reuse. The main requirements on grant holders are to:

50 See: <http://www.nerc.ac.uk/about/access/statement.asp>

51 NERC, *Data policy handbook*, (2002), available at: <http://www.nerc.ac.uk/research/sites/data/documents/datahandbook.pdf>

- follow the principles in the data policy handbook and be aware of the role of the designated data centres in order to liaise with them as appropriate;
- prepare a written data management plan before a project is started;
- create data to comply with the minimum accepted standards in terms of completeness and documentation;
- offer a copy of any dataset resulting from NERC-funded activities to its data centres together with documentation/metadata describing these data so they can be made available for future research;
- individual scientists, principal investigator teams and programmes will be permitted a reasonable period of exclusive access to datasets which they have collected, allowing them to work on them and produce publications;
- to cooperate in validating and publishing data in their entirety - when this can be justified in terms of their scientific value - rather than merely creaming off a subset for immediate publication in the literature;

Support provided

- Guidance for researchers on data creation and management is provided through the specialist data centres.⁵²
- To improve access to published outputs, NERC has established a centralised e-print research repository.⁵³
- The NERC network of data centres supports the long-term curation of its environmental data holdings. Access to all NERC funded data is facilitated through an integrated, searchable catalogue.

3.2.7 STFC - Science and Technology Facilities Council⁵⁴

A statement on access to published research outputs is available within STFC's *Research Grants Handbook*.⁵⁵ As yet, no formal data policy has been produced by the Council, but a degree of curation continues informally through channels established by CCLRC and PPARC.

Policy stipulations

Overview

- Researchers are expected to make publications that stem from STFC-funded research available at the earliest opportunity. No time limits are specified for data.
- The STFC does not appear to require a statement on data management or sharing. Plans that formalise ownership and agree distribution mechanisms were required by PPARC at the proposal stage, but it is unclear to what extent this is still required.
- Researchers are expected to make published outputs accessible. Data produced by STFC-funded researchers is often made available, however a formal requirement to do so does not appear to be in place.

52 For details see: <http://www.nerc.ac.uk/research/sites/data/>

53 See: <http://www.nerc.ac.uk/about/access/repository.asp>

54 Please note that STFC was formed in April 2007 as a merger of CCLRC & PPARC

55 STFC, *Research Grants Handbook*, section 8.2: dissemination, (2008) available at: <http://www.scitech.ac.uk/rg/rghDisplay2.aspx?m=s&s=64>

- The Council does not have a formal policy covering long-term curation, however the recently released strategy consultation document notes it is a priority to continue maintaining appropriate data stores, showing a commitment to sustained access.⁵⁶
- It does not seem compliance with the STFC's open access policy will be monitored.

Published outputs

STFC supports the RCUK position on access to research outputs. As such, it recommends institutions include publication fees as indirect costs. In addition, it is required of STFC-supported researchers that:

- the full text of articles resulting from STFC grants that are published in journals or conference proceedings must be deposited, at the earliest opportunity, in an appropriate e-print repository, wherever such a repository is available;
- the bibliographical metadata (including a link to the publisher's website) must wherever possible be deposited, at or around the time of publication, in the relevant e-print repository.

These requirements are subject to compliance with publisher's copyright and licensing policies. STFC runs an e-Publications archive to help researchers make articles, conference papers, technical reports, theses and books available.

Data

Although a formal data policy has not been developed following the merger of CCLRC and PPARC, it appears some curation continues with decisions being made on a project-by-project basis. Research data are currently made available through a variety of subject specific data centres and portals, such as the UK Solar System Data Centre, the Chemical Database Service and the Diamond Data Portal.⁵⁷

Support provided

- Best practice guides for researchers creating and maintaining digital material were not found on the STFC website, though some assistance may be provided informally through the data centres.
- An e-Publications archive is run by the Council to collate and make its published research accessible.⁵⁸
- Several data centres and research portals are in place for unpublished research and data, such as the UK Solar System Data Centre, the Chemical Database Service and the Diamond Data Portal. These are generally organised on a subject basis rather than serving the outputs of the whole council, and it seems deposit is ad hoc as opposed to a condition of grant funding.

56 STFC, *Consultation document on strategy*, Section 2.4.1.6, p44 available at:

<http://www.stfc.ac.uk/stfcconsultation/sources/strategy/StrategyConsultationDocument.pdf>

57 For additional details see: <http://www.stfc.ac.uk/ResFac/Data.aspx>

58 See: <http://epubs.stfc.ac.uk/index>

3.2.8 The Wellcome Trust

The Wellcome Trust has both a statement on open access of research outputs,⁵⁹ and a concise *Policy on data management and sharing*.⁶⁰ A very thorough Q&A on how to comply with the data policy is also provided.⁶¹ The Trust's *Guidelines on Good Research Practice* provide details of the standards of good research conduct expected of funded researchers.⁶²

Policy stipulations

Overview

- Certain time limits are stipulated for both publications and research data. Published outputs should be deposited as soon as possible, and in any event within six months of final publication. As an absolute minimum, researchers should make relevant data available to others on publication of their research (providing this is consistent with any ethics approvals, consents and intellectual rights) however opportunities for timely and responsible pre-publication sharing of data should also be maximised. Research institutions are also required to maintain data securely for a minimum of ten years.
- In cases where a resource of benefit to the research community is planned, or a significant quantity of data that could potentially be shared is likely to be generated, researchers are required to submit a data management and sharing plan. The guidance indicates that these plans should, where relevant to the proposed research, take account of five areas: data quality and standards; use of public data repositories; intellectual property; protection of research participants; and long-term preservation and sustainability.
- The Wellcome Trust actively promotes open access to research publications resulting from its funded research. The Trust provides additional funding to institutions to cover open access charges and requires researchers make publications available through PubMed Central and UK PubMed Central. Researchers are also expected to maximise the availability of research data with as few restrictions as possible.
- The Trust's *Guidelines on Good Research Practice* encourage appropriate creation and maintenance of research resources. Institutions are expected to have guidelines setting out responsibilities and procedures for the appropriate storage and disposal of data and samples.
- Compliance with the open access policy on published research outputs is monitored, as is access to major resources. The Trust does not actively monitor data sharing on grants across the board.

59 Wellcome Trust, *Position statement in support of open and unrestricted access to published research*, (2008), available at: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm>

60 Wellcome Trust, *Policy on data management and sharing*, (2007), available at: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

61 Q&A: *Wellcome Trust policy on data management and sharing*, available at: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-management-and-sharing/WTX035045.htm>

62 Wellcome Trust, *Guidelines on Good Research Practice*, (2005), available at: <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002753.htm>

Published outputs

The Wellcome Trust is committed to maximising the benefit generated by the research it supports. As such it expects the researchers it funds to make publications openly accessible. The Trust will provide additional funding to its grant holders through their institutions, to cover open access charges. The Wellcome Trust:

- requires electronic copies of any research papers that have been accepted for publication in a peer-reviewed journal, supported in whole or in part by Wellcome Trust funding, to be made available through PubMed Central and UK PubMed Central as soon as possible, and in any event within six months of the journal publisher's official date of final publication;
- encourages - and where it pays an open access fee, requires - authors and publishers to license research papers such that they may be freely copied and reused (for example for text and data-mining purposes), provided that such uses are fully attributed.

Data

The Trust believes benefits gained from research data will be maximised when they are made widely available to the research community as soon as feasible, so that they can be verified, built upon and used to advance knowledge. Its policy on data management and sharing states its expectation that its funded research should maximise the availability of research data with as few restrictions as possible.

For grant proposals intending to develop a community resource or in which a significant quantity of data that could be shared will be generated, the Trust requires researchers to submit a data management and sharing plan as part of their application.

In addition, as outlined in its *Guidelines on Good Research Practice*, the Trust expects:

- clarity at the outset of the research programme as to the ownership of data and samples used or created in the course of the research and the results;
- researchers to keep clear and accurate records of the procedures followed and the approvals granted during the research process;
- data generated in the course of research to be kept securely in paper or electronic format, as appropriate - the Trust considers a minimum of ten years to be an appropriate period, but research based on clinical samples or relating to public health might require longer storage;
- institutions to have guidelines setting out responsibilities and procedures for the storage and disposal of data and samples (including compliance with the requirements of any ethics committee);
- back-up records to be kept for data stored on a computer.

Support provided

- The Wellcome Trust has provided an excellent guide on how to comply with its data policy, but does not appear to have made many other sources of guidance or best practice available to assist researchers in creating high quality and reusable data.
- A repository for published research outputs is provided through UK PubMed Central.
- Responsibility for curation falls to the institutions in which Wellcome Trust funded researchers are based. As yet data curation facilities are not provided, however the Trust will meet curation costs as outlined in the data management and sharing plan. The Wellcome Trust also supports the European Bioinformatics Institute to ensure bioinformatics data remain accessible through community databases.

4. Policies related to digital curation

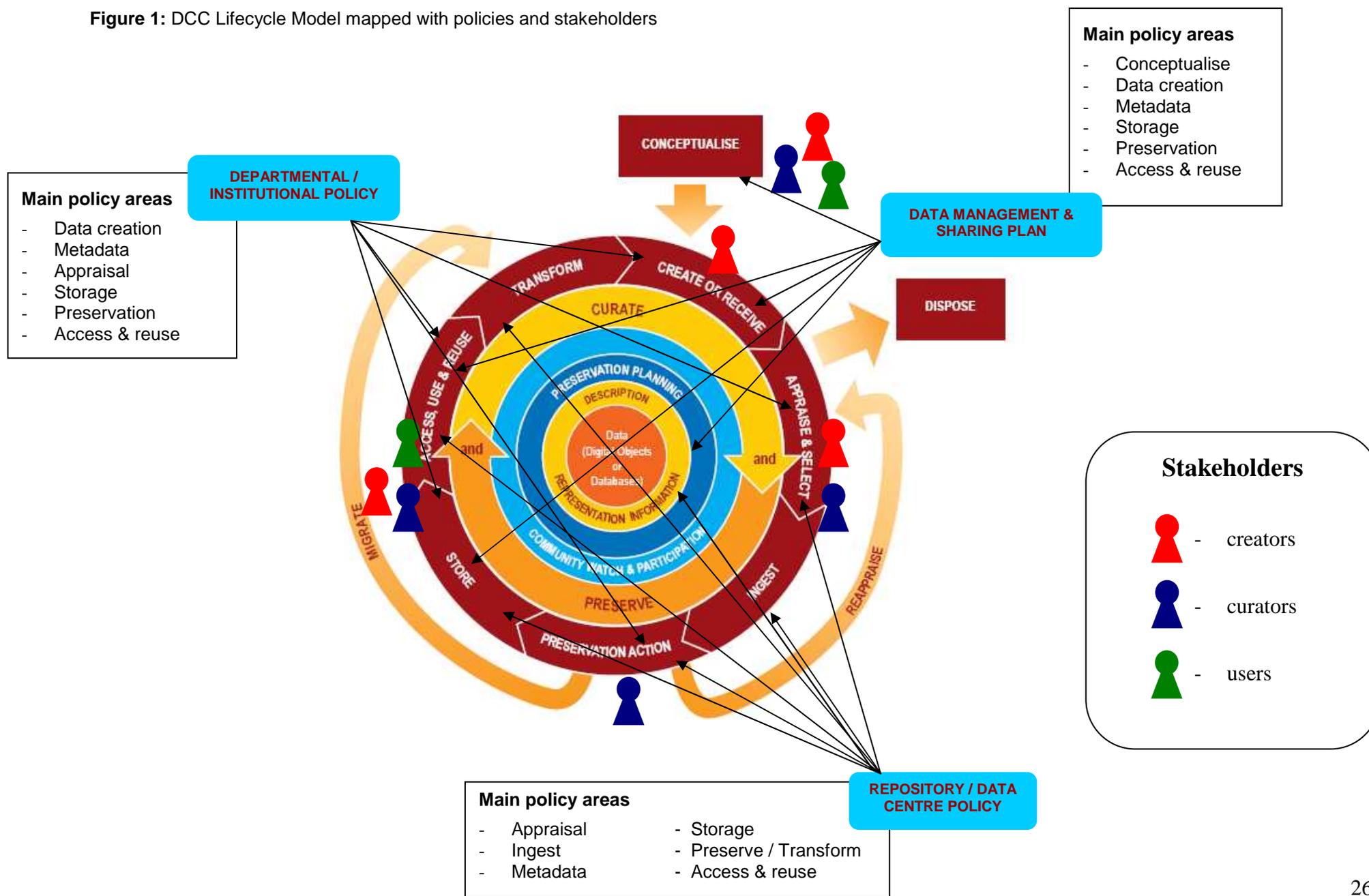
4.1 Overview

The curation policies of funding bodies essentially require two things: first that research outputs are created in an appropriate manner to ensure they can be made widely accessible; and second that they are maintained in the long-term to facilitate continued access, either under the auspices of the institution in which the funded-researcher is based or by means of deposit in a specialist repository or data centre. Maintaining digital resources throughout their entire lifecycle involves the input of several stakeholders and requires a variety of policies. A model of the inputs is shown overleaf in the annotated DCC Curation Lifecycle Model.

As the diagram shows, both the roles of various stakeholders and the coverage of curation policies significantly overlap. The main stakeholders used in the model are creators, curators and users. It should be noted that these roles are not discrete: creators for example may also take on responsibility for long-term curation in cases where a data centre is not available. Some activities require input from multiple stakeholders, such as the initial conceptualisation phase and appraisal and selection. It is likely a content creator would consult with curation experts and the potential user community when developing a proposal, in order to ensure the outputs are created in an appropriate way to meet user requirements.

The policies noted on the model are naturally not exhaustive, yet do cover the main ones that should be in place to ensure research outputs are effectively curated throughout their lifecycle. There is significant overlap in the content of each policy, so anyone developing a new policy will need to be mindful of how it fits in with existing policies to ensure they work in harmony and that policy stipulations aren't conflicting. Again this calls for collaboration between each stakeholder group. A good example of this working in practice is in the Research Councils that provide comprehensive curation support to researchers. NERC data centre staff, for example, assist with data management plans at the application stage. As such, they can ensure the proposed creation and maintenance of research outputs meets NERC's own policies and standards, as well as those of the data centres that will be responsible for the resources in the long-term. This approach ensures smooth delivery throughout the curation lifecycle and promotes good relations between all of the stakeholders, thereby minimising risk to the data and ensuring their continued access and use.

Figure 1: DCC Lifecycle Model mapped with policies and stakeholders



4.2 Current policy provision

4.2.1 Repository & data centre policies

Of the three policies outlined on the model overleaf, the most prevalent are Institutional Repository or data centres policies. Most of these organisations will have some form of curation policy in place. These will vary in comprehensiveness from basic statements on what can be deposited and how this will be stored and accessed, to very detailed descriptions of processes and procedures followed to ensure long-term preservation. The OpenDOAR directory of open access repositories provides details of each repository's metadata, reuse, content, submission and preservation policies, where available.⁶³ The service has also created an online policy tool with example clauses and varied levels of openness that repository managers can select from to create their own policies.⁶⁴ Data centres often have very detailed curation policies. An example that could be used for best practice is that of the UKDA.⁶⁵ This defines roles and responsibilities and details activities that will take place during various stages of the curation lifecycle.

Assistance for repositories wanting to develop policies is available through OpenDOAR, the Repositories Support Project⁶⁶ and SUETr, a training programme for start up and enhancement repository projects.⁶⁷ Complementary services have also been established to help researchers comply with funder, publisher and repository mandates. The SHERPA JULIET service gives details of funding body expectations on open access publishing and archiving, while the related RoMEO service provides a database of publishers' copyright and self-archiving policies.⁶⁸ The OpenDOAR directory can then be used to identify a suitable repository for deposit that complies with the relevant funder and publisher mandates.

4.2.2 Data management and sharing plans

Curation policies related to specific research projects are on the increase. The majority of the main UK research funders expect applicants to consider creation and management of their research outputs at the proposal stage in order to submit a data management and sharing plan. These generally require researchers to cover what data and related documentation will be created, how this will be achieved i.e. methodologies and standards to be used, how the data will be maintained and access and reuse potential, bearing in mind any IPR and ethical considerations. In many cases there is still a need to join up data management plans with related institutional and repository preservation policies to which the research outputs will be subject. At present this only really occurs in instances where a supporting curation infrastructure is provided, for example in the case of NERC and ESRC, as data centre staff can liaise with researchers from the proposal stage.

Support for researchers preparing data management plans is urgently required, and the DCC is working to fill this gap by providing policy templates and a research grant mediation and support service. A template data management plan based on the requirements of the main UK research funders will be provided in Spring 2009. Guidance is also available through other curation centres such as the UKDA, which provides extensive details on aspects to consider when managing and sharing data,⁶⁹ and has provided tailored data management support for the cross-council Rural and Economic Land Usage (RELU) programme.⁷⁰ Australian and US guidance could also provide useful references to UK researchers. The Australian National

63 See: <http://www.opendoar.org/index.html>

64 The OpenDOAR policy tool is available at: <http://www.opendoar.org/tools/en/policies.php>

65 Woollard Matthew, *UK Data Archive Preservation Policy*, v3.0, (2008), available at: <http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf>

66 See RSP website at: <http://www.rsp.ac.uk/repos/rules>

67 See the SUETr wiki at: <http://suetr-rp.wikispaces.com/>

68 For the JULIET service see: <http://www.sherpa.ac.uk/juliet/index.php> and for RoMEO: <http://www.sherpa.ac.uk/romeo/>

69 See the guidance pages at: <http://www.data-archive.ac.uk/sharing/sharing.asp>

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

University has released a valuable data management manual, which provides guidance on various procedures such as file transfer, version control, backup, security and data documentation.⁷¹ A section of the manual provides guidance on writing data management plans. Very useful advice and a data planning checklist are also provided by MIT libraries.⁷²

4.2.3 Institutional curation policies

The area requiring most work in terms of developing curation policy is at an institutional level. Although most institutions have records management policies, several recent reports have noted the lack on institutional policies for digital curation more generally and have noted that development of these is sporadic.⁷³ The need for an institutional stance on curation was also noted in audits conducted by the Data Audit Framework project.⁷⁴ During interviews several researchers requested institution level guidance with which to frame their work and to guide development of departmental data policies or data management plans. The RCUK consultation on a code and policy for the governance of good research conduct⁷⁵ may well heighten the urgency for institutions to implement curation policies, particularly if the guidance is changed to reflect the Australian code⁷⁶ which requires institutions to have a policy on retention, ownership, and access to data to be eligible to receive research funding.⁷⁷

The Beagrie policy study attempts to effect change in this area by providing a framework for those planning to create institutional preservation policies, together with a series of mappings to core university business drivers to embed the policies in wider aims. The study found particular synergies between preservation policies and information, library and records management strategies. A basic guide on creating a preservation policy has also been created by JISC, which provides an overview for those at the outset of policy development.⁷⁸

If an institutional stance is not provided, a gap will remain in the overarching policy framework that jeopardises the longevity of digital research outputs. A suite of curation policies that interact and complement one another is needed. To develop curation policies and put them into practice, all stakeholders will require significant help and guidance. The DCC has started to collate and produce relevant tools and guidance to support those creating curation policies.

71 Australian National University, *ANU Data Management Manual: Managing Digital Research Data at the Australian National University*, v1.03, (2008), available at:

http://ilp.anu.edu.au/dm/ANU_DM_Manual_v1.03.pdf

72 See: <http://libraries.mit.edu/guides/subjects/data-management/>

73 See for example, Lyon, Liz, *Dealing with Data*, Beagrie, Neil et al, *Digital preservation policies study*, and Ruusalepp, Raivo, *Comparative study of international approaches to sharing research data*

74 Interviews were held in which researchers were asked about their working practices and data issues.

To find out more see the project website at: <http://www.data-audit.eu/>

75 RCUK, *Code of Conduct and Policy on the Governance of Good Research Conduct*, (2008), available at: <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/reviews/grc/consultation.pdf>

76 *Australian Code for the Responsible Conduct of Research*, (2007), available at:

<http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>

77 Discussion of the Australian code is provided on the DCC curation blog and DataShare blogs at: <http://digitalcuration.blogspot.com/2008/09/national-data-mandate-australian-code.html> and <http://jisc-datashare.blogspot.com/2009/02/data-walkabout-5-brisbane-queensland.html>

78 See: <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/establishing-a-digital-preservation-policy/>

4.3 Aspects to consider when creating curation policies

There are various aspects that researchers and institutions should consider when creating curation policies. New policies are likely to require a change in working practices, which can be difficult to achieve so particular attention should be paid to fitting policies to existing workflows and structures. Consideration should also be made of any additional infrastructure or support services that will be required for implementation, so these can be put in place at the same time. Above all policies need to have a clear purpose and be explicit about the activities to be undertaken and staff responsibility for these. Initial scoping surveys and a phased roll out could be helpful to ensure the policy is appropriate and to secure buy-in. Particular aspects to consider are noted below in more detail.

Purpose and scope

What is the rationale for the policy and expected benefits? This may link back to the mission statement of your organisation. Parameters should also be defined with regard to the type of research outputs covered by the policy, the curation activities being addressed, and the context in which the policy is to be applied i.e. across an entire institution or just a single department or research project. Enforcing minimum standards of data quality and developing robust appraisal decisions will be crucial for cost-effectiveness.

Roles and responsibilities

A policy should define who will undertake the work and assign responsibility. By connecting policies with current activities, proposed tasks can be aligned with existing workflows and staff skills. When assigning new responsibilities, changes should be thought through to ensure they are practical and implementable. Defining roles will also be important for accountability.

Examples / best practice

It is worth checking if there are example policies or relevant frameworks you could use as a guide. The recent JISC digital preservation policies study has an annotated bibliography, which points to several existing curation policies that could be of assistance.⁷⁹ In addition the DCC is creating template curation policies based on funder requirements and accepted best practice, that users can download and customise.

Implementation

Policies need to be accompanied by the appropriate services, skills and infrastructure if their implementation is to be feasible. The wider context should also be considered: how does this policy fit in with and complement existing policies? A scoping exercise could be used as a precursor to identify needs for the policy and raise awareness of proposed changes. Phasing the implementation could be useful to test applicability and encourage staff buy-in.

Review

The curation environment is rapidly changing. Technological developments, cultural shifts and legislative change are all altering the working practices of content creators, curators and users. Policies need to be regularly reviewed and adjusted to ensure they remain appropriate.

⁷⁹ Beagrie, Neil et al, *Digital preservation policies study*, (2008), pp37-44, available at: http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf

4.4 Summary and recommendations

All areas of curation policy require further development, with the most growth called for in institutional policies. Some tools and services are available to support policy development, particularly in the repository sector. As yet though, there has not been sufficient time for stakeholders to respond to the requirements funding bodies have issued. There are some signs of change: for example DCC staff are currently involved in a policy scoping study at the University of Glasgow, which will identify staff needs for curation support and provide recommendations to develop an institution-wide policy.⁸⁰ The DCC expects the coming years to see other such policy developments and is providing tools and services in support of this.

Recommendations for groups planning to develop curation policies are:

- Resources such as the OpenDOAR policy tool, UKDA and DCC guidance should be used where possible to help close gaps in the curation policy landscape. Making policies available will also help others build on best practice.
- Policies need to be mindful of context: a data management plan for example needs to complement and work in harmony with the relevant institutional and repository requirements for curation.
- Existing structures could be used to embed curation in research workflows, for example researchers could be directed to advice on data management as part of funding application procedures in the same way ethical approval is currently ensured.
- Existing staff, such as subject librarians, Freedom of Information officers or departmental representatives could take on a broader support role to act as curation champions and broker relations between staff and the various support services.
- Attention should be paid to encouraging uptake of any new policy. Preliminary scoping exercises, test phases or a reward system that recognises researchers who adopt best practice could be useful.
- Policies need to be practical and accompanied by the required support infrastructure to ensure they can be implemented. A mechanism to monitor implementation and revise the policy to amend inappropriate clauses is crucial.

There are also three broader issues to be addressed collaboratively by all stakeholders if we are to create a stable base from which to develop meaningful curation policies:

1. Identify roles and responsibilities for curation;
2. Assess costs and benefits to determine how and by whom curation should be financed;
3. Develop a robust and sustainable curation infrastructure with appropriately skilled staff.

Research has been started in all these areas over the past few years, most notably by the JISC.⁸¹ The DCC has been addressing skills gaps through its DC101 training course and the research data management forum.⁸² It is crucial we continue to build on this initial research.

80 For details of the study see: <http://www.hatii.arts.gla.ac.uk/research/prespolicy.html> The lessons will be shared through the DCC to provide an example developing an institutional preservation policy

81 See for example: Swan, Alma, *Skills, role and career structure of data scientists and curators*, (2008), available at: <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf>, Fry, Jenny et al, *Identifying benefits arising from the curation and open sharing of research data*, (2008), available at: http://ie-repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf and the repositories programmes: <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005.aspx> and <http://www.jisc.ac.uk/whatwedo/programmes/reppres>

82 For RDMF on skills see: <http://www.dcc.ac.uk/events/data-forum-2008-november/>

5. Conclusion

There have been a number of promising policy developments in recent years. The value of research outputs and potential to enhance knowledge through wide and open access to these has been recognised by funding bodies and is reflected in their policies. Moreover advances in technology facilitate international data sharing and enable new modes of research. Although initial steps have been taken by research funders to promote digital curation, there is much work still to be done if researchers and institutions are to meet these requirements.

Roles and responsibilities need to be defined in order that appropriate training and support can be provided. At present, curation activities fall to several groups, including researchers, libraries, information professionals, data centres and repositories. There is often a lack of clarity as to who is responsible in each case and insufficient communication between the groups. How curation should be funded is also an open question, yet to be resolved. More research into the benefits of curation will be crucial to ensure cost-effectiveness.

There are significant gaps in the curation infrastructure which undermine policy requirements. These are being addressed, for example by the JISC digital repositories programmes. The urgency for institutions to develop a support infrastructure and curation policies is growing. Complementary policies also need to be developed by other stakeholders involved in the curation lifecycle, such as researchers, departments and repositories. A considered approach is needed to ensure these policies work in harmony. It's not yet clear how best to develop and implement policies at local level. Lessons will need to be drawn from early exemplars.

While there is still much work to be done in the area of curation policy, the prospects for safeguarding our research outputs are good. Most funding bodies are pushing for open access and data sharing. In order to create and maintain our research outputs for future generations we must be clear what role each stakeholder plays, how the respective activities fit together, and ensure there are no gaps in the curation lifecycle. The policies we developing and putting into practice at present will help us achieve this.

6. Bibliography

Reports

- Beagrie, Neil et al, *Digital preservation policies study*, (2008), http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf
- Gibbs, Harry et al, *DISC-UK DataShare: State-of-the-Art Review*, (2007), available at: <http://www.disc-uk.org/docs/state-of-the-art-review.pdf>
- Lyon, Liz, *Dealing with Data: Roles, Rights, Responsibilities and Relationships*, (2007), http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf
- RIN, *Research funders' policies for the management of information outputs*, (2007), <http://www.rin.ac.uk/files/Funders%27%20Policy%20&%20Practice%20-%20Final%20Report.pdf>
- RIN, *Stewardship of digital research data: a framework of principles and guidelines*, (2008), <http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20ofull%20version%20-%20final.pdf>
- Ruusalepp, Raivo, *A comparative study of international approaches to enabling the sharing of research data*, (2008), available at: http://www.dcc.ac.uk/docs/publications/reports/Data_Sharing_Report.pdf

Policies⁸³

- Australian Code for the Responsible Conduct of Research, (2007), <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>
- BBSRC, *Data Sharing Policy*, (2007), http://www.bbsrc.ac.uk/publications/policy/data_sharing_policy.pdf
- European Science Foundation, *Good Scientific Practice in Research and Scholarship*, Policy Briefing 10 (2000), <http://www.esf.org/publications/policy-briefings.html>
- MRC, *Policy on data sharing and preservation*, <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm>
- MRC, *Principles for access to, and use of, MRC funded research data*, (2007) <http://www.mrc.ac.uk/consumption/groups/public/documents/content/mrc003759.pdf>
- NERC, *Data policy handbook*, (2002), <http://www.nerc.ac.uk/research/sites/data/documents/datahandbook.pdf>
- OECD, *Principles and Guidelines for Access to Research Data from Public Funding*, (2007), <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

⁸³ Note the AHRC and ESRC do not have separate policy documents, their requirements are provided in the respective research funding guides

- RCUK, *Code of Conduct and Policy on the Governance of Good Research Conduct*, (2008), <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/reviews/grc/consultation.pdf>
- UK Data Archive, *Preservation Policy*, v3.0, (2008), <http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf>
- Wellcome Trust, *Policy on data management and sharing*, (2007), <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>

Open access statements

- AHRC, *Guidance on access to research outputs*, <http://www.ahrc.ac.uk/FundingOpportunities/Documents/access%20to%20research%20outputs.pdf>
- BBSRC, *Policy on access to research outputs*, (2008), http://www.bbsrc.ac.uk/publications/policy/access_research_outputs.html
- EPSRC, *Policy on access to research outputs*, <http://www.epsrc.ac.uk/AboutEPSRC/AccessInfo/ROAccess.htm>
- ESRC, *Guidance on access to research outputs*, <http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Support/access/>
- MRC, *Position statement in support of open and unrestricted access to published research*, (2008), <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Openaccesspublishing/Positionstatement/MRC004368>
- NERC, *Guidance on access to research outputs*, <http://www.nerc.ac.uk/about/access/statement.asp>
- RCUK, *Position statement on access to research outputs*, (2006) <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf>
- STFC, open access statement in *Research Grants Handbook*, section 8.2, (2008) <http://www.scitech.ac.uk/rgh/rghDisplay2.aspx?m=s&s=64>
- Wellcome Trust, *Position statement in support of open and unrestricted access to published research*, (2008) <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002766.htm>

Research funder guidance

- AHDS, *Guides to Good Practice, Information Papers and Case Studies*, <http://www.ahds.ac.uk/about/publications/index.htm>
- AHDS, *Preservation handbooks*, <http://ahds.ac.uk/preservation/ahds-preservation-documents.htm>
- BBSRC, *Statement on Safeguarding Good Scientific Practice*, http://www.bbsrc.ac.uk/publications/policy/good_scientific_practice.pdf

- EPSRC, *Guide to Good Practice in Science and Engineering Research*, (2006), <http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/GoodPracticeGuideSciEngRes.pdf>
- Economic and Social Data Service, Guidance on creating data, data management and deposit: <http://www.esds.ac.uk/aandp/create/createintro.asp>
- MRC, *Good Research Practice Guide*, (2000), <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002415>
- UKDA, Data management and sharing web pages: <http://www.data-archive.ac.uk/sharing/sharing.asp>
- Wellcome Trust, *Guidelines on Good Research Practice*, (2005), <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002753.htm>
- Wellcome Trust, Q&A: *Wellcome Trust policy on data management and sharing*, <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-management-and-sharing/WTX035045.htm>

Related projects and websites

- Australian National University data management pages: <http://ilp.anu.edu.au/dm/>
Particularly see the manual at: http://ilp.anu.edu.au/dm/ANU_DM_Manual_v1.03.pdf
- Glasgow University policy study: <http://www.hatii.arts.gla.ac.uk/research/prespolicy.html>
- MIT data management advice: <http://libraries.mit.edu/guides/subjects/data-management/>
- OpenDOAR directory of open access repositories: <http://www.opendoar.org/index.html>
- OpenDOAR policy tool: <http://www.opendoar.org/tools/en/policies.php>
- JISC digital media preservation policy development advice: <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/establishing-a-digital-preservation-policy/>
- JULIET registry of research funders' open access policies: <http://www.sherpa.ac.uk/juliet/index.php>
- RELU programme guidance on data management: <http://www.data-archive.ac.uk/relu/>
- Repositories Support Project: <http://www.rsp.ac.uk/>
- RoMEO registry of publisher copyright policies & self-archiving: <http://www.sherpa.ac.uk/romeo/>
- Start-up and Enhancement Repositories Training Programme (SUETr): http://www.ukoln.ac.uk/repositories/digirep/index/SUE_Training_Programme:_SUETr and <http://suetr-rp.wikispaces.com/>
- StORE (Source to Output Repositories) project: <http://www.jisc.ac.uk/whatwedo/programmes/digitalrepositories2005/store.aspx>