



## **DIGITAL CURATION POLICY FRAMEWORK**

Prepared by Graham Pryor  
Associate Director, Digital Curation Centre

July 2010

## **PART 1 – Introduction and rationale**

This policy framework comes in two parts:

- Part 1 explains how the framework has been developed, its scope and relevant tools and standards;
- Part 2 presents the nine-part curation policy framework, with explanatory text and suggested components for an institutional policy document.

### **1.1 Origin and purpose**

This document has been produced by the ERIS Project Work-package 2 team. ERIS, a JISC-funded initiative, seeks to develop a set of user-led and user-centric solutions that will motivate researchers to deposit their work in repositories and facilitate the integration of repositories into research and institutional processes. As a result, it is envisaged that a trusted cross-repository and resource discovery service will be delivered, which will be capable of providing access to a critical mass of Scottish research output.

The ERIS team envisages that the production of a Curation Policy Framework will enable the promulgation of institution-wide curation policies informed by established methods and standards. Use of the Framework, which would underpin the needs of established or emerging institutional repositories, should ensure an evenness of approach across the Scottish HEI community that is consistent with support to an environment of collaborative and cross-institutional research (as exemplified in Scotland by the Research Pools). The Framework has been designed in the wake of an investigative survey of the status of digital curation and preservation policies in Scottish HEIs between November 2009 and January 2010,<sup>1</sup> which found that

- repositories are still relatively young and the need to apply explicit curation policies is only beginning to be acknowledged;
- no institution level preservation policies exist, although repository level policies were identified at four institutions;
- there is a very low level of awareness of both existing preservation policies and digital preservation issues amongst administrative and research staff.

It may be deduced from this analysis that the majority of Scottish HEIs are at a similar stage in addressing the need for implementing curation policies, and as a consequence they can be expected to experience a similitude of challenges and opportunities. Given, therefore, that this is almost universally a formative period it does provide some advantages of simplification from building a generic framework on a green field site.

### **1.2 Key issues from the survey**

The authors of the survey report identified the following key components of a curation policy framework for HEIs:

- policies should cover all academic output, including research data and other digital materials;
- a risk based analysis of research outputs should be employed to establish the requirements for preservation policies at an institutional level;
- institutional policies should include the clear assignment of roles and responsibilities;
- the profile of preservation and curation processes, and therefore the value of an institutional data management policy, will be enhanced where repository services actively provide advice on preservation, curation, data policy and compliance.

---

<sup>1</sup> The report from the survey can be found via the Specialist Documentation link at <http://sligachan.lib.ed.ac.uk/wordpress-mu/project-outputs/>.

These points have been incorporated in the scope for the Curation Policy Framework described in Part 2.

### 1.3 Scope of framework

This framework has been developed on the assumption that it will be applied to all managed academic output in digital form within an institution, where the expectation is that institutional data management services are principally but not exclusively supplied by the offices of an institutional repository.

Our survey found that within HEIs the responsibility for digital academic output is divided between different departments and services, which suggests that the development of policies could also be fragmented and according to the diverse views of individual departments. We recommend therefore that the ownership of curation policies should be delegated to the institutional repository service in order to achieve consistency and sustainability. Repositories are themselves governed by established principles, which should routinely include a strategic programme for preservation planning and action.<sup>2</sup>

The framework addresses the needs of a curation policy at institutional level, which differs in certain respects from a preservation policy. The DCC has also produced a preservation policy framework for repositories.<sup>3</sup> The distinction between preservation and curation is made in a DCC Briefing paper from 2006<sup>4</sup> that describes **digital preservation** as

the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organisational considerations that relate to the survival and management of digital material.

As noted in the paper, the biggest risk to digital objects is the continual change to the hardware and software environment, which can often have a detrimental effect on the authenticity and integrity of a preserved resource. Hence the need for **digital curation**,

which builds upon the underlying concepts of digital preservation whilst emphasising opportunities for added value and knowledge through the provision of annotation and continuing resource management.

Preservation is therefore a significant part of curation, but curation implies the more active and ongoing management of a preserved digital resource. It will include a cycle of reappraisal, transformation (to enable new uses) and disposal, migration to more accessible or durable formats, and a steady focus on the changing needs of data users. Part 2 of this document may therefore be referred to alternatively as a data management policy framework.

### 1.4 Established approaches to preservation and curation strategy

In developing this framework, reference has been made to a number of extant policy documents, including the UK Data Archive Preservation Policy (<http://www.data-archive.ac.uk/news/publications/preservationpolicy.pdf>), the Arts and Humanities Data Service Collections Preservation Policy (<http://ahds.ac.uk/documents/colls-policy-preservation-v1.pdf>) and the DISC-UK Datashare guide to Policy-making for Research Data in Repositories (<http://www.disc-uk.org/docs/guide.pdf>). Creation of the framework has also been enabled by resources available from the Digital Curation Centre, not least the series of steps described by the Curation Lifecycle Model that link curation activities with the

---

<sup>2</sup> The EC-funded PLATTER Repository Planning Checklist and Guidance recommends the Ten Core Principles of Trust Repository Design, developed in 2007 by The Center for Research Libraries (CRL), The Digital Curation Centre (DCC), Digital Preservation Europe (DPE) and The German Network of Expertise in Digital long-term preservation (nestor) – see <http://www.digitalpreservationeurope.eu/platter.pdf>, page 9.

<sup>3</sup> DCC, *Preservation policy template for repositories*, (2010) available at: <http://www.dcc.ac.uk/sites/default/files/Preservation%20policy%20template.pdf>

<sup>4</sup> Continued Access to Authentic Digital Assets, Maureen Pennock, <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/digital-preservation>.

continuum of data creation, use and re-use (<http://www.dcc.ac.uk/resources/curation-lifecycle-model>).

### 1.5 Relevant tools and standards

As a first principle, and assuming that the institutional repository is the acknowledged custodian of institutional curation policy, it is recommended that its goals, objectives and performance targets are defined according to a mechanism such as the Planning Tool for Trusted Electronic Repositories (PLATTER),<sup>5</sup> which will contribute to its achieving trusted status amongst its stakeholders. **PLATTER** is designed to complement existing audit and certification tools by providing a framework that will allow new repositories to incorporate the goal of achieving trust into their planning from an early stage. PLATTER's ten core principles for a trusted repository include a having a commitment to the continuing maintenance of digital objects for identified communities, a strategic programme for preservation planning and action, and a technical infrastructure adequate to continuing maintenance and security of its digital objects, all of which are essential to effective digital curation.

Building on one of the key outcomes from the survey, where it was found that the requirements for data curation would be linked to the governance of risk, it is further recommended that institutions schedule and conduct periodic **DRAMBORA**<sup>6</sup> audits to evaluate the repository's ability to fulfil its self-specified goals. DRAMBORA is an audit methodology based on risk assessment; available online, it guides users through the audit process on a step-by-step basis, and provides export functionality which can feed in to management reporting systems.

Risk governance implies the management of valuable institutional assets, of which research data comprise a significant proportion of the intellectual asset base. Continuing on that theme, the **DATA ASSET FRAMEWORK**<sup>7</sup> (DAF) helps ensure that research data are preserved and remain accessible in the long term by providing a means to identify, locate and describe research data within an institution, leading to an assessment of how they are being managed. The DAF audit process can be applied to identify weaknesses in data policy and current data creation and curation procedures, and can be used in both the planning of a new policy and in testing the efficacy of one that is operational.

As explained in 2.1, below, an especially critical aspect of the framework is the development of routines for data management planning, which should be pervasive across all research undertakings. Here, the DCC's **CHECKLIST FOR A DATA MANAGEMENT PLAN**<sup>8</sup> provides an invaluable reference work covering the broad scope of issues to be addressed.

### 1.6 A practical approach to implementing a curation policy

The value of any policy is to be found in the extent to which it is respected and adopted, as well as from the practical advantages enjoyed by those who follow it. Within the large and heterogeneous communities that make up a university, where it is rarely appropriate or rewarding to direct all activity according to a rigid, top-down business regimen, a one-size-fits-all policy is unlikely to attract universal probity. Indeed, in this context caution should be exercised even in the use of the term *policy*, which can be viewed adversely in a community accustomed to the close performance of critical analysis, especially if it is coupled with the potentially arcane concept of *data curation*. A 'policy' statement by senior management, recognisably a necessary means of legitimising and communicating the institution's clear commitment to good practice, might therefore be described instead as a protocol, a memorandum of principle or a statement of intent, with the purpose of gaining attention rather than exercising assertion.

---

<sup>5</sup> PLATTER, *Op Cit*.

<sup>6</sup> <http://www.dcc.ac.uk/resources/tools-and-applications/drambora>.

<sup>7</sup> <http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework>.

<sup>8</sup> A link to the checklist is provided from <http://www.dcc.ac.uk/resources/policy-and-legal/data-management-plans>  
DMP Online, a versatile web-based version of the checklist, is available in beta form at <http://dmponline.hatii.arts.gla.ac.uk>

## DRAFT version 1.12

Consequently, and as a first step, the approach taken in this advisory framework focuses on the initial building blocks necessary for the creation of a culture in which effective data management (and hence data curation) is as routinely pursued as financial, safety or other essential good working practice. Only when these building blocks have become established is it likely that the climate will be more receptive to formalisation into a more regulatory declaration of policy.

The initial statement used by senior management to launch this pre-policy initiative will be limited to introducing the rationale for institutional attention to good data management, and will quickly focus on the tools and services to be provided, as well as an explanation of the actions and responses expected from the research community. Hence, the 'policy' will recognise examples of existing good practice, introduce the concept and method of data management planning as a means of exporting coherence across the institution, describe the responsibilities of colleges, schools and departments, and outline the training and support services that will be provided.

Hence the institutional policy will in its first iteration consist of an action plan necessary for achieving specific goals. Recommended goals are the creation and registration of data management plans at appropriate levels within the institution, the development of a training programme for specific cohorts of researchers, and the establishment of a help and advisory service. Each of these is a practical deliverable for which arguments of authenticity and value may be presented. A data management planning checklist is provided as an appendix to this document.

Longer term, when data management planning has become embedded as a research 'literacy', established good practice may be formalised into a structured set of principles that define the requirements, responsibilities and actions necessary to support the fully-fledged curation of digital academic output. These undertakings will reflect the working procedures that will already be commonplace in both national and institutional repositories. Only at this juncture may a pan-University policy be more confidently described and promulgated.

For this second phase, nine separate components are considered appropriate to a policy framework. They are described in Part 2, where explanatory text is provided in italics and model policy content is given in normal script within frames.

## PART 2 – A curation policy framework

This curation policy framework consists of nine component parts. It presents both the conditions under which the policy is being introduced and the practical measures necessary for compliance with the policy. Thus, at 2.1 through 2.3, the policy is given its official endorsement and a definition of responsibilities and expectations, followed from 2.4 with the practical measures that should be included to satisfy policy requirements.

Under each heading the text in italics explains the points that should be considered, while the text in the boxes provides example content for an institutional policy.

Please note that the phrase 'researchers and/or data custodians' has been used to describe those who are directly responsible for data management and should be amended to reflect local institutional arrangements.

### **2.1 Policy statement**

*An unequivocal statement, preferably issued under the signature of a member of senior institutional management, which describes the purpose and validity of the curation policy, drawing attention to the breadth of its jurisdiction (to whom it applies and to which digital objects), its provenance, the consequences from non-compliance, and key contacts for advice and support.*

Considerable attention has been given recently to the management of publicly funded scholarly output, initially concerning the provision of open access to published research but more currently in terms of enabling access to research data (i.e. the materials from which scholarly publications are generated). A desire to extract maximum value from investment in research has led a number of the key funding agencies to produce policies that include a requirement for grant applicants to explain their arrangements for access and sharing, management and the long-term curation of their research outputs. Compliance with acceptable arrangements is expected increasingly to represent an obligation upon research grant holders. This institution recognises these trends and obligations; it also recognises that the data produced by its researchers can comprise valuable assets requiring effective custodianship and management. This data curation policy is intended to assist the University in meeting its obligations to funders, and therefore protecting its eligibility as a grant applicant, whilst also providing a mechanism for assuring the safekeeping of a major asset base.

This policy for the management of scholarly output in digital form applies to all data upon which research outputs are (or are likely to be) published, as well as the published outputs themselves. The policy will apply to publications and data managed in an institutional repository serving the university community and to publications and data managed by or within individual research units.

Assistance will be provided in meeting the terms of this policy, in the form of help and guidance by repository staff, including online reference material, as well as through the provision of training by experts in digital curation.

Specific questions arising from the content of this policy should be addressed in the first instance to [*name/email/telephone no. of repository/other staff*]. All other issues should be directed to senior management via the appropriate departmental fora.

### **2.2 Data audit**

*It is recommended that audits of data assets are conducted at a departmental or research centre level (or at other equivalent levels beneath those of college or school) to ascertain the volume and condition of academic output in digital form. As a result of these audits, it should be possible to make judgements as to the specific requirements for curation (storage, back-up, security, access, data management planning, and the provision of curation expertise).*

Academic departments [*insert other unit names as appropriate*] are required to audit the research data collections for which they are the responsible producers and keepers, in order to establish what data are held, by whom, their size, format and condition. The arrangements made for storage, and procedures for management, access, sharing and disposal should be assessed against a scale of risk from corruption or loss, inadequate accessibility, under-utilisation and reuse, and the cost of inappropriate retention. Data audits should be recurrent within a three to five year cycle.

Information gathered from the audits, when recorded in a departmental or institutional registry of data assets, should be augmented by decisions that classify sets of data according to their predicted value (usefulness) in the short, medium and long term. This in turn will inform decisions as to their ongoing curation needs.

Advice on the conduct of data audits will be provided by institutional repository staff. A useful data audit methodology is explained on the Digital Curation Centre's website at <http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework>.

### 2.3 Roles and responsibilities

*A description of the responsibilities of data producers, data owners and data custodians, including an explanation of the remit for curation assigned to central services (e.g. the institutional repository service) and with contacts identified for policy, service and training.*

Roles and responsibilities for effective data curation are remitted to two broad groups: data creators and data managers. Data creators may be individual researchers or research teams; data creators may also be data managers. At an institutional level, data managers may be the staff of the repository or other university-wide data service.

It is the responsibility of **data creators** to develop a data management plan at the conceptual stage of any new research programme and to embed that plan within the research process. It is recommended that plans are linked directly with the research lifecycle.<sup>9</sup> A central registry of data management plans will be maintained by the University's repository service [*or Research Support Service or alternative; could be part of the institutional Research Information System, where these exist*].

**Principal Investigators** are responsible for ensuring that data produced are managed according to the approach defined in the data management plan for their research programme.

**Data managers** operating at a **research team** or departmental level are required to follow the processes of accepted good practice that are described in more detail in the following sections, which will include the provision of metadata and documentation to a minimum specified standard, data security arrangements that include storage and back-up procedures, clearly specified protocols for access and reuse of data, and measures to ensure compliance with legislative, ethical and organisational expectations.

**Data managers** acting for the **institutional repository** will be responsible for the selection, ingest and curation of digital objects supplied for long-term preservation, as well as the provision of a help service offering guidance and advice on data management planning and techniques for electronic data management. Guidance for the process of depositing research data will be posted on the repository website, giving clear instructions on timing and method. A registry for recording the results of data audits is also maintained by the institutional repository [*or Research Support Service or...*].

Human Resources [*or the Research Support Service or...*] are responsible for the provision of training in basic data management skills to research staff and for the embedding of such training in the induction programmes for new research postgraduates.

<sup>9</sup> A preferred approach may be found in the Digital Curation Centre's CURATION LIFECYCLE MODEL, which can be downloaded from <http://www.dcc.ac.uk/resources/curation-lifecycle-model>.

## 2.4 Procedures for implementation and operation

*Briefly describe, using as little jargon or technical language as possible, the key processes/workflow for ensuring effective data curation, drawing on the phases defined in the DCC Curation Lifecycle Model, and showing timings, actions and agents (inclusion of a diagram of the Model is recommended as a visual aide). Make sure the availability of help and guidance is made clear for any of the activities described, which may contain unfamiliar concepts.*

Researchers and/or data custodians are expected to apply the following routines:

- Planning: describe under the following three sub-headings how data will be 1) created, 2) captured and 3) stored, referring to the methods, equipment, software and environment to be used.
- Define which *metadata* will be used (administrative, descriptive, structural and/or technical metadata) and describe how it is to be assigned. Guidance on metadata provision will be available from the institutional repository [*or Research Support Service or...*].
- Describe the process for evaluating and selecting data for long-term curation and preservation, with reference to documented guidance, policies or legal requirements.
- Document the routines for the transfer of data to the institutional or other repository or custodian.
- Describe the conditions and actions that will be undertaken following transfer to ensure the long-term preservation and retention of data, including data cleaning, validation and the rules/protocols governing access, use and reuse.

## 2.5 Security and integrity

*Introduce by reminding individuals that digital information can easily be copied, altered or deleted and that, where data are being curated outside of an institutional repository, measures to ensure data security and integrity must still be applied. The actual processes will therefore be presented in a manner that is applicable to both departmental and institutional data stores. Cross-reference to the institution's standard IT security/back-up/disaster recovery guidance and procedures should be inserted as appropriate.*

Because digital information can so easily be copied, altered or deleted it is important at any time to be able to demonstrate its authenticity and to prevent unauthorised access. The basic steps to be taken should include:

- Areas and rooms designated for data storage must be fit for purpose, being structurally sound and free from the risk of flood and fire, and with appropriate access control.
- Keep a master file of data under the control of a nominated individual.
- Provide restricted access only to master versions of data by requiring written or online authentication and the maintenance of access records.
- A formal procedure for the destruction of master files.
- Recorded changes to master files.
- Routine and periodic back-up of all curated data, in formats suitable to long term preservation, supplemented by the archiving of master files in a secure repository. Guidance on preferred data formats for long term software readability is available from the institutional repository [*or Research Support Service or...*].



## 2.6 Review, refreshment and transformation

*The preparation of data for long-term preservation and retention as an authentic, reliable and usable resource involves such actions as data cleaning, validation, assigning metadata and ensuring acceptable data structures or file formats. These should be periodically reviewed, tested and sustained. Reviews also provide an opportunity to add significant value to data by providing additional metadata to broaden their possible applications in new and different avenues of research. Finding new uses for data will only remain a viable option, however, if steps are taken to sustain its accessibility and usability and discoverability? (slightly different to accessibility, although may be beyond the remit of this Framework).*

To ensure that data remains both accessible and usable, researchers and/or data custodians should:

- Copy or migrate data files to new media between two and five years after they were first created.
- Undertake sample searches using the assigned metadata.
- Add metadata that reflect standard parameters and/or generic derived variables in other

## 2.7 Version control

*It is important that different copies or versions of data held in different formats or locations are subject to version control. Principally, this means that, where appropriate, changes to one version of a dataset will be reflected in other versions. Where data files continue to be used (as opposed to being archived) it is important to keep track of the version that contains the most current data, to know who has changed what aspect of the data, and to record the reason for any change.*

Researchers and/or data custodians are expected to:

- Maintain master files of data in a suitable format (see 2.5 above) to identify and protect the authoritative version.
- Identify datasets uniquely using a systematic naming convention (guidance is available from the institutional repository [*or Research Support Service or...*]).
- Clearly record the version and status of all datasets (e.g. draft, interim, final, internal).

## 2.8 Disposal and withdrawal

*Data should be retained only where there is a justifiable reason. The less data retained, the lower the cost of managing it and the fewer the risks. Some data must be retained to comply with legal or business requirements. Disposal must be undertaken in a regulated manner in order to protect it from misuse, particularly where there are issues of confidentiality or intellectual property or other potential sensitivities.*

Researchers and/or data custodians should:

- Dispose of data that has not been selected for long-term curation and preservation in accordance with documented policies, guidance or legal requirements. Guidance on data (selection and?) retention policies is available from the data protection officer and the records management or institutional repository service [*insert as appropriate per institution*].
- Transfer data to an archive, repository, data centre or other custodian in accordance with written procedures. The institutional repository service will advise.
- Data selected for destruction must be destroyed according to secure procedures advised by the institutional IT services [*insert as appropriate per institution*].

## **2.9 Sustainability**

*Introduce the concept of sustainability by posing questions describing potential threats to the continuity of the curated resource, such as -*

- *What will happen if the repository is closed or funding is reduced?*
  - *Has sufficient funding been predicted to support a growth in data volume and use?*
  - *If university funding is reduced will data services be cut or have ongoing costs and a long-term strategy been planned for?*
  - *Are plans in place to transfer repository content elsewhere if necessary?*
- and follow by providing a series of actions that resolve the main issues of sustainability.*

In the event of the institutional repository service being withdrawn, as a consequence for example of funding constraints or changes in strategic priorities

- data will be transferred to another appropriate archive (see below)
- items will be returned to their originators upon request

Agreements will be reached with other institutional or domain-specific repositories to provide succession arrangements in the event that internal procedures become untenable. Such agreements have already been concluded with [*add list*]. Of these, for example, and serving the social sciences domain, the UK Data Archive (UKDA) is committed to supporting continued funding for all of its operations relating to preservation management, and makes every effort to remain up-to-date with any relevant technological advances to ensure continued access to its collections.

Should internal reorganisation alter the strategic accountabilities for data management, the University Executive Committee [*or Senior Management Team or...*] will be responsible to the governing body for ensuring that appropriate replacement arrangements are made.

Arrangements for the effective data management/curation of scholarly output will be included in the institutional disaster management [*or recovery or...*] plan.

**This framework is being tested in a series of case studies at four Scottish institutions during 4Q/2010. Subsequently, it will be revised to reflect the outcome of those studies.**

## **DRAFT version 1.12**

**Appendix – Checklist for a data management plan**

**To be added.**