

The University of Edinburgh

Central Management Group

20 April 2011

Policy for Management of Research Data

Brief description of the paper

This paper outlines the proposed policy for management of research data. The policy aims to help the University and researchers implement the UK Research Integrity Office's Code of Practice requirements for collection, management, security and retention of research data, prioritising appropriate infrastructure, systems, services and training. Adoption of this policy will improve the University's services for researchers, increase its ability to protect valuable research outputs, assist it to meet legal and ethical obligations and make more effective use of IT resources to support research.

The policy was drafted by a working group which included representatives from all Colleges, Information Services, ERI, and records management. The policy has been available widely for consultation, in particular to College library and computing committees. It has been approved by Library Committee, Knowledge Strategy Committee, and Research Policy Group.

It is acknowledged that this is an aspirational policy, and that implementation will take some years. Implementation plans for both research data management and research data storage are in development, and will address issues which have been raised during the consultation period.

Action requested

For approval and onward submission to Court.

Resource implications

Does the paper have resource implications? If accepted, there will be resource implications for the running of individual research projects to best practice and the provision of central services for the University. It is anticipated that some of the additional resources required can be recouped from research funders.

Risk assessment

Does the paper include a risk analysis? Yes. An outline risk analysis is included.

Equality and diversity

Does the paper have equality and diversity implications? No

Freedom of information

Can this paper be included in open business? Yes

Originators of the paper

- Chris Rusbridge, Consultant
- Sheila Cannell, Convenor of the Research Data Management Working Group
- Peter Clarke, Convenor of the Research Data Storage Working Group
- Members of the Research Data Management and Research Data Storage Working Groups

Paper presented by

Jeff Haywood, Vice Principal Knowledge Management, CIO and University Librarian

Policy for Management of Research Data

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.
9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

Background

1. Introduction

Data in one form or another are critical for most research, both as primary inputs and first-order outputs. At international, national and local levels, there is intense interest in how to manage the rapidly expanding volume and complexity of research data. Concern is both for the shorter term – ensuring competitive advantage through secure and easy-to-use access, and for the longer term – ensuring enduring access and usability to the research community into the future and compliance with legislation. The UK government and research funding bodies are debating with the HE community how best to address this large and complex problem, and have funded various initiatives to explore options (including data archives such as the UK Data Archive¹, the proposed United Kingdom Research Data Service², and the Edinburgh-based Digital Curation Centre³).

There is already much good practice in the University of Edinburgh, with many Principal Investigators (PIs) already developing data management plans. This policy is to ensure consistency of practice across the University, acknowledging the variations in disciplines across the University.

Most Research Councils now mandate or encourage Data Management Plans and deposit of data for later re-use where practical⁴.

The Research Information Network (RIN) has published a framework of principles and guidelines for the stewardship of digital research data⁵. The RIN Framework derives from prior work by the OECD⁶.

The UK Research Integrity Office (UKRIO) has prepared a standard Code of Practice for Research⁷ that is regularly reviewed to take into account changes in legislation, and to reflect national and international best practice. The University of Edinburgh has formally adopted the UKRIO Code of Practice as its own policy. The Code demands the highest standards of researchers, but also requires the University to set up systems, procedures and infrastructure to support them properly.

Scholarly journals in increasing numbers are requiring that continuing access to underlying data sets⁸ be provided by first or corresponding authors.

The JISC Support of Research Committee⁹ has various programmes dealing with research data, the latest being the JISC Managing Research Data Programme (JISCMRD)¹⁰.

The UK Research Data Service (UKRDS) project¹¹ started with the objective of assessing the feasibility and costs of developing and maintaining a national shared digital research data service for UK Higher Education sector. The project sponsors saw this as forming a crucial component of the UK's e-infrastructure for research and innovation, which would add significantly to the UK's global competitiveness.

The UKRDS feasibility study concluded that embedding the skills, capability and organisation into the HE research management process was the best approach and that a relatively small national service structure (likely to be developed from the Edinburgh-based Digital Curation Centre) would be needed to foster this through channelling training, tools and good practice developed by existing national and international skill centres.

¹ <http://www.data-archive.ac.uk/>

² <http://ukrds.ac.uk/> and Final Report at <http://ukrds.ac.uk/resources/>

³ <http://www.dcc.ac.uk/>

⁴ <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

⁵ <http://www.rin.ac.uk/data-principles>

⁶ OECD. (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding. Paris: OECD. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

⁷ <http://www.ukrio.org/resources/UKRIO%20Code%20of%20Practice%20for%20Research.pdf>

⁸ Eg <http://www.nature.com/nature/journal/v458/>. See also Baggerly, K. A. (2010). Disclose all data in publications. *Nature*, 467(401), 60-60.

⁹ <http://www.jisc.ac.uk/aboutus/committees/subcommittees/jsr.aspx> This programme seeks to expand effective data management and data sharing to benefit research and the HE sector more generally. The JISC is working towards developing a national strategy with key stakeholders (Research councils, Funding Councils, Institutions etc.), in order to help to establish the foundations for an effective UK research data infrastructure.

¹⁰ <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

¹¹ <http://www.ukrds.ac.uk/>

What is clear is that there will be no external solution that will remove from the University the requirement to provide storage and management procedures for the data resulting from its own research activities.

At the University of Edinburgh, a consultation on computing requirements of the research community has been conducted¹². Key findings of this consultation indicated a need for larger storage space on servers; more robust archiving services; simple, secure and preferably automatic data back-up services; and a high demand for training and awareness raising across the University.

Second, a pilot implementation of the JISC Data Audit Framework project¹³ was carried out. The study focused primarily on research data management rather than storage requirements. The findings at Edinburgh were that there was inadequate storage space and lack of clarity about roles and responsibility for research data management by University research staff. The project noted a need for storage and backup procedures including provision for business continuity arrangements. A formal procedure was needed for data transfer when staff and students leave the institution.

Two projects have recently been initiated. The internally funded project LAIRD¹⁴ aims to help build two-way links between research articles and the data that support findings reported in the article. A new JISC-funded training project MANTRA¹⁵ aims to develop online learning materials which reflect best practice in research data management, grounded in three disciplinary contexts: social science, clinical psychology, and geoscience.

Solutions for the University of Edinburgh will only be successful if they come from a partnership of individual researchers, Schools, Colleges and Information Services. Each has expertise and resources that can be brought to bear to the benefit of all.

Last year, the Research Computing Advisory Group (RCAG) consulted with a representative sample of staff and research students and produced a strategy plus implementation roadmap¹⁶, which recommended to the Vice Principal that addressing research data storage and management was a high priority requirement.

The oversight of research computing has now been made the responsibility of the re-instated IT Committee, and as part of its 2009-10 Work plan, it is taking up the challenge of producing a review of data storage and management, starting its work with **research data** (leaving learning and teaching data and corporate data to a later date). Two groups have been set up, with close links:

- (i) Research Data Storage Working Group
- (ii) Research Data Management Working Group.

This paper draws on many different inputs, including comments and documents from these Working Groups, the authors' own experiences, the Digital Curation Centre (DCC), the Data Audit Framework Project Steering Committee's recommendations¹⁷, the ERIS project guidelines for data policies, the UK Research Integrity Office Code of Practice for Research, and other University policies. The paper brings these together in a draft.

¹² Cuna Ekmekcioglu. Research Computing Consultation Report. December 2007. Available from Edinburgh DataShare repository.

¹³ Cuna Ekmekcioglu, Robin Rice. Edinburgh Data Audit Implementation Project. January 2009: <http://ie-repository.jisc.ac.uk/283/>. See also Data Audit Framework: <http://www.data-audit.eu/index.html> (now referred to as the Data Asset Framework)

¹⁴ See <http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/laird>

¹⁵ See <http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/mantra>

¹⁶ http://www.rcg.isg.ed.ac.uk/docs/open/ResearchStrategy_May08_public2.pdf

¹⁷ See project page at <http://www.ed.ac.uk/schools-departments/information-services/about/organisation/edl/data-library-projects/edinburgh-data-audit>

2. Constraints

Sharing and access to research data are encouraged or required by explicit policies of several Research Funders¹⁸, and increasingly the policies of Editorial Boards also require the retention and sharing of data substantiating research articles¹⁹.

The Freedom of Information (Scotland) Act and the Environmental Information Regulations (Scotland) provide legal force to the public's right to know, subject to particular exemptions or exceptions. The Data Protection Act controls access to personal data and the Data Protection Principles provide a framework for personal data processing. The Freedom of Information (Scotland) Act includes an exemption for ongoing programmes of research; this exemption is not available in the rest of the UK. Paradoxically, a policy that research data will be published at some future date may provide an exemption in the rest of the UK (which may become an issue in research collaborations).

Implementation of the European INSPIRE Directive²⁰ in the UK will require increasing sharing of geospatial datasets created by public authorities such as the University.

The UKRIO Code of Practice states: "Organisations and researchers should ensure that research data relating to publications is available for discussion with other researchers, subject to any existing agreements on confidentiality (13.12.1). Data should be kept intact for any legally specified period and otherwise for three years at least, subject to any legal, ethical or other requirements, from the end of the project. It should be kept in a form that would enable retrieval by a third party, subject to limitations imposed by legislation and general principles of confidentiality (13.12.2)." The Code further points out: "Organisations and researchers working with, for, or under the auspices of, any of the UK Departments of Health and/or the National Health Service must adhere to all relevant [data management] guidelines, for example the Department of Health's *Research Governance Framework for Health and Social Care*²¹.

The Code also places responsibilities with the University. For research data specifically "*Organisations* should have in place procedures, resources (including physical space) and administrative support to assist researchers in the accurate and efficient collection of data and its storage in a secure and accessible form. (3.12.5)"

University policy constraints include Records Management Policies, and IT Security and other IS policies. Ethics Committees may place particular requirements on certain research data, and these must be fulfilled.

Some of these constraints may prevent data being retained or deposited, while some may allow deposit but limit or control sharing or the terms and conditions for sharing. However, it is important to note that these constraints apply throughout the research lifecycle, not just to research data outputs.

¹⁸ See DCC Cross Council Policy Overview at <http://www.dcc.ac.uk/resources/policy-and-legal/cross-council-policy-overview>

¹⁹ For example, see Nature's policy, described in Nature (2009). Authorship policies. *Nature*, 458(7242), 1078. doi: 10.1038/4581078a. Retrieved from <http://www.nature.com/nature/journal/v458/n7242/index.html#ed>

²⁰ See http://www.agi.org.uk/storage/inspire/inspire_intro.pdf

²¹ See

http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4122427.pdf

3. Policy aims and objectives

The aims of this proposed policy are to:

- Support the University's mission for "the creation, dissemination and curation of knowledge".
- Support research excellence.
- Help the University and Researchers implement the UK Research Integrity Office's Code of Practice requirements for collection, management, security and retention of research data, prioritising appropriate infrastructure, systems, services and training.
- Protect the legitimate interests of the University, of research data subjects and of other parties.
- Acknowledge differing practices in different disciplines.
- Support appropriate openness and transparency, and ensure accountability for the use of public funds.

4. Policy

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data which is retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.
9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

5. Roles and responsibilities

There are many stakeholders involved in the management of research data and the implementation of this policy. Practice is and will remain very different in different research domains and sub-domains. Wide involvement by stakeholders, for example, PIs, research funders, the University, will be needed to achieve this policy's aims.

Since this policy is about research excellence, the policy aims to ensure that responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with PIs.

PIs are responsible to the University (through their research group, School or College) for the management of research data arising from their research projects or programmes throughout the life of those projects or programmes (working data). Where research is conducted with other institutions, Edinburgh Co-PIs or Co-Investigators are responsible for management of research data under their control and held by the University. Management of research data includes management of all metadata, documentation and software/hardware resources required to properly manage and analyse the data.

PIs are responsible for offering data that substantiate their research findings for deposit in an appropriate domain or University repository (and should register with the University details of data deposited to an external service). PIs may offer other data with potential re-use value from their research project for deposit in an appropriate repository.

The University, through its various structures (College, School, research group or institute, Administration and Information Services) continues to be responsible for supporting PIs and their researchers as far as possible, through policies, services, systems and infrastructure, training, support and advice. This will assure confidentiality, security and integrity of data including research data (eg backup systems applicable to research data on all applicable platforms). [Services to support researchers could make it easier to transfer data to University control at an appropriate point.

The University (IS) should be responsible for any research data deposited in University repositories, and associated systems and services. Working data remain the responsibility of the PI.

The University (Records Management) is responsible for advice on records retention, including research records, which may include research data.

Research Funders (UK and international) increasingly require PIs to create Research Data Management Plans or protocols at project proposal stage, and to take responsibility for the implementation and maintenance of these plans throughout their research project. The University should be responsible for assisting researchers in creating Research Data Management Plans (including recouping data management and curation costs where possible), and capturing proposed Plans as records. Research Data Management Plans should cover a broad range of issues including data capture, management, integrity, confidentiality and security, data ownership, sharing and publication, and data deposit, retention and/or destruction.

6. Research data assets and research data management plans

Research Funders in the UK are increasingly requiring Data Management Plans with research proposals. All new and proposed research projects should create research data management plans²². The University should support PIs in the creation and implementation of these plans. A data planning checklist is available at <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/data-planning-checklist>, and advice on data management plans is at <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/data-mgmt-plan>.

²² See DCC Data Management Plan content checklist at http://www.dcc.ac.uk/sites/default/files/documents/tools/dmpOnline/DMP_checklist_v2.2_100106-publicVersion.doc

There should be a section on research data management in all ethics applications, covering data confidentiality, security and integrity issues.

Schools should from time to time assess the data assets²³ associated with research in which members of the School participate. Data assets should be interpreted broadly to include data together with documentation and metadata for use, and may include specialist software.

The Research Data Storage Working Group is recommending improved University, College and/or School services in support of research data, not least to help maintain its security, integrity and confidentiality where required²⁴.

7. Data ownership and control

In many cases factors including the collaborative and international basis of many research projects make the nature and extent of intellectual property rights in research data unclear. Any assertion of intellectual property rights to data should be made clear at the outset of any research project and should explicitly form part of any collaboration or partnership agreement and Data Management Plan. Paradoxically, asserting rights to data can be an important element in protecting data in cases where confidentiality is required, or even in making data open. Similarly, in the UK beyond Scotland, a commitment to publish research data in the future may be enough to provide exemption to premature disclosure. Cross-border collaborative research agreements should consider such issues.

Data ownership and control are about much more than intellectual property rights, however. Data ownership implies stewardship and good management of the data. Ensuring data remain accessible is an important part of that stewardship, and depositing data in a repository can pass on that responsibility to others capable of discharging the responsibility over a longer period. Indeed, transferring some data into University custody while still retaining ownership and control could be valuable.

The University urges researchers to make their data open once research is published²⁵ or after an agreed embargo period. Open Data approaches reduce the cost of FoI and other requests for access and re-use, are compatible with accountability and openness, and encourage the re-use that maximises the benefit to society from publicly-funded research. These approaches cannot be used in all cases, however, for a variety of reasons, including ethics, and privacy and exploitation of intellectual property, and reduced or restricted access to data are acceptable where these apply.

Where researchers seek to make their research products open, explicit devices such as the Creative Commons Attribution²⁶ licence (for Creative Works such as text and multi-media documents) should be used. For data, a licence such as the Creative Commons [CC0 waiver](http://creativecommons.org/licenses/by/2.5/scotland/)²⁷ or the Open Data Commons Public Domain Dedication and Licence ([ODC-PDDL](http://www.opendatacommons.org/odc-public-domain-dedication-and-licence/)²⁸) should be used. These licences will make the situation clear to potential re-users; the absence of a licence may mean resources are not re-used and hence do not get cited. Licences with a “Non-commercial” restriction may seem attractive but should be avoided where possible, as they severely limit re-usability (the interpretation of non-commercial being unclear).

Where data are deposited in a University repository, the University does not require transfer of ownership of the data to the University. However, ownership transfer may simplify the long-term

²³ Ekmekcioglu, C., Rice, R., Jerrome, N., Breeze, J., Grace, S., Knight, G., et al. (2009). Data Asset Framework: Implementation Guide. Edinburgh. Retrieved from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

²⁴ A separate study by RDSWG makes recommendations for services relating to storage of research data [reference to be added]

²⁵ Eg the Panton Principles appear to be attracting support, see at <http://pantonprinciples.org/>

²⁶ Known as CC-BY, see <http://creativecommons.org/licenses/by/2.5/scotland/> This licence allows others to copy, distribute, display, and perform the work, and to make derivative works, but the original author must be given credit when re-using the work.

²⁷ See <http://wiki.creativecommons.org/CCZero>

²⁸ See <http://www.opendatacommons.org/odc-public-domain-dedication-and-licence/>

management of data. A non-exclusive licence to hold, manage and preserve the data²⁹ is essential and a non-exclusive licence to make the data available is highly desirable.

Where an external or domain repository or service takes a data deposit, the repository usually holds the data under a deposit agreement that governs their (non-exclusive) rights both to hold the data and to make the data available under certain conditions. The repository usually does not own any rights to the data; any rights remain with the original owners.

8. Risk

If appropriate policies are not followed, the university will be exposed to potential reputational and financial risk. See for example the UEA Climate Research Unit³⁰ and QUB tree ring³¹ cases recently, and the press release from JISC on related topics³².

Failure to establish appropriate policies will mean the University will breach its own adopted UKRIO Code of Practice.

Failure to provide scalable services for research data will lead to individual highly variable practice, some of which will fall well short of excellence, and which in aggregate may greatly exceed the cost of scalable solutions.

Failure to make adequate research data and documentation available for analysis and verification may lead to the University being the subject of unwelcome articles such as Baggerley and Coombes (2009³³).

²⁹ The Edinburgh DataShare deposit agreement is at <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/data-repository/depositor-agreement>

³⁰ See UEA Media and Communications special site at <http://www.uea.ac.uk/mac/comm/media/press/CRUstatements>

³¹ See Smith, G. (2010). Freedom of Information Act 2000 (Section 50) Environmental Information Regulations 2004 Decision Notice [Queen's University Belfast]. Wilmslow. Retrieved from http://www.ico.gov.uk/upload/documents/decisionnotices/2010/fs_50163282.pdf .

³² See <http://www.jisc.ac.uk/news/stories/2010/07/opendata.aspx>

³³ Baggerly, K. A., & Coombes, K. R. (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 3(4), 1309-1334. doi: 10.1214/09-AOAS291.