

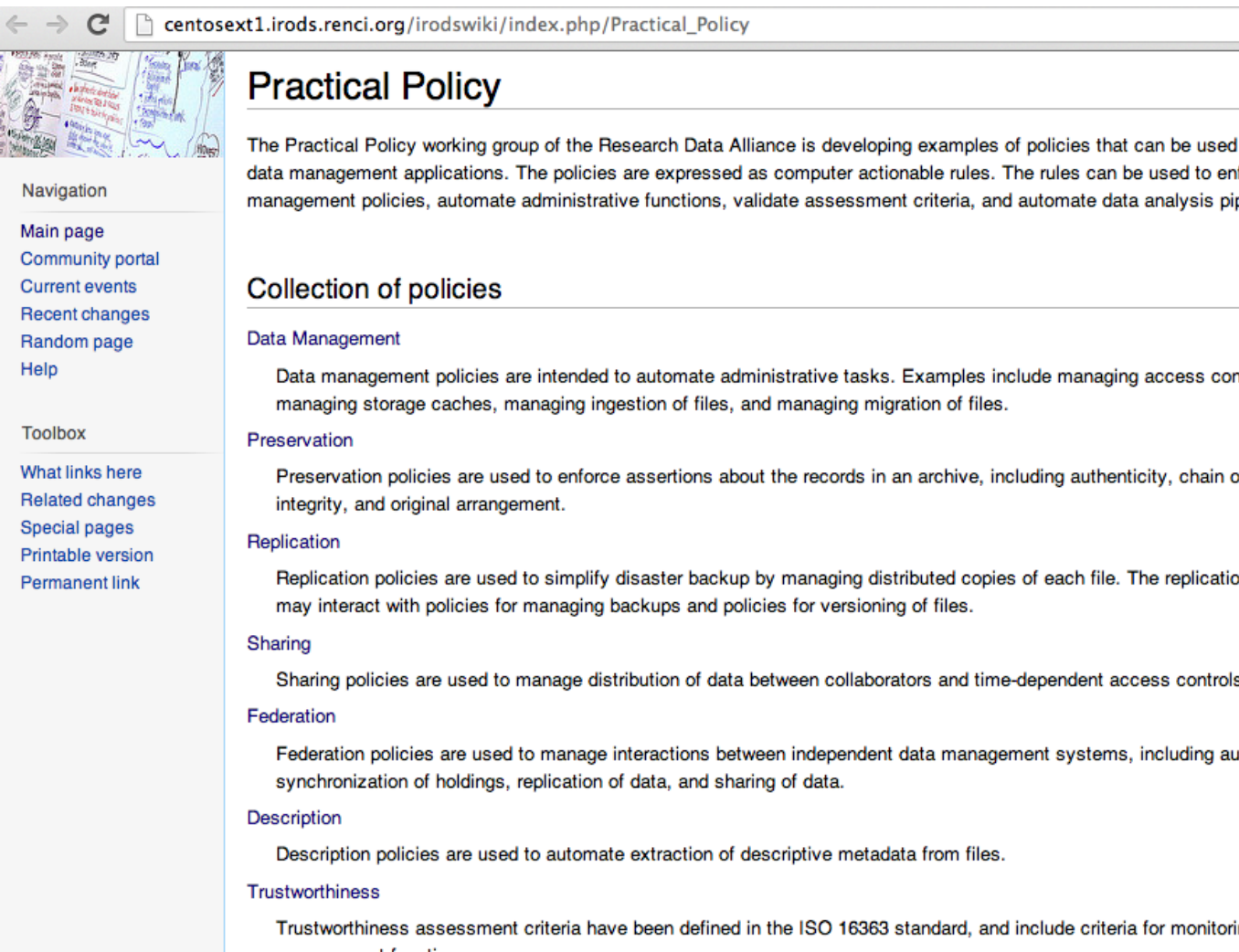


Research Data Policies

Seachange or Zeitgeist?

PERSPECTIVES ON RESEARCH DATA

What the OAI audience might think...



The screenshot shows a web browser window with the URL `centosext1.irods.renci.org/irodswiki/index.php/Practical_Policy`. The page title is "Practical Policy". The main content area is titled "Practical Policy" and contains the text: "The Practical Policy working group of the Research Data Alliance is developing examples of policies that can be used in data management applications. The policies are expressed as computer actionable rules. The rules can be used to enforce management policies, automate administrative functions, validate assessment criteria, and automate data analysis pipelines." Below this, there is a section titled "Collection of policies" which lists several categories: "Data Management", "Preservation", "Replication", "Sharing", "Federation", "Description", and "Trustworthiness". Each category has a brief description of the policies it covers. For example, "Data Management" policies are intended to automate administrative tasks, and "Preservation" policies are used to enforce assertions about the records in an archive.

Navigation

- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

Toolbox

- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

Practical Policy

The Practical Policy working group of the Research Data Alliance is developing examples of policies that can be used in data management applications. The policies are expressed as computer actionable rules. The rules can be used to enforce management policies, automate administrative functions, validate assessment criteria, and automate data analysis pipelines.

Collection of policies

Data Management

Data management policies are intended to automate administrative tasks. Examples include managing access control, managing storage caches, managing ingestion of files, and managing migration of files.

Preservation

Preservation policies are used to enforce assertions about the records in an archive, including authenticity, chain of custody, integrity, and original arrangement.

Replication

Replication policies are used to simplify disaster backup by managing distributed copies of each file. The replication policies may interact with policies for managing backups and policies for versioning of files.

Sharing

Sharing policies are used to manage distribution of data between collaborators and time-dependent access controls.

Federation

Federation policies are used to manage interactions between independent data management systems, including authentication, synchronization of holdings, replication of data, and sharing of data.

Description

Description policies are used to automate extraction of descriptive metadata from files.

Trustworthiness

Trustworthiness assessment criteria have been defined in the ISO 16363 standard, and include criteria for monitoring and assessing the trustworthiness of data.

Data Management

Data management policies are used to automate administrative tasks. Examples include managing access control, managing storage caches, managing ingestion of files, and managing migration of files.

Preservation

Preservation policies are used to enforce assertions about the records in an archive, including authenticity, chain of custody, integrity, and original arrangement.

Replication

Replication policies are used to simplify disaster backup by managing distributed copies of each file. The replication policies may interact with policies for managing backups and policies for versioning of files.

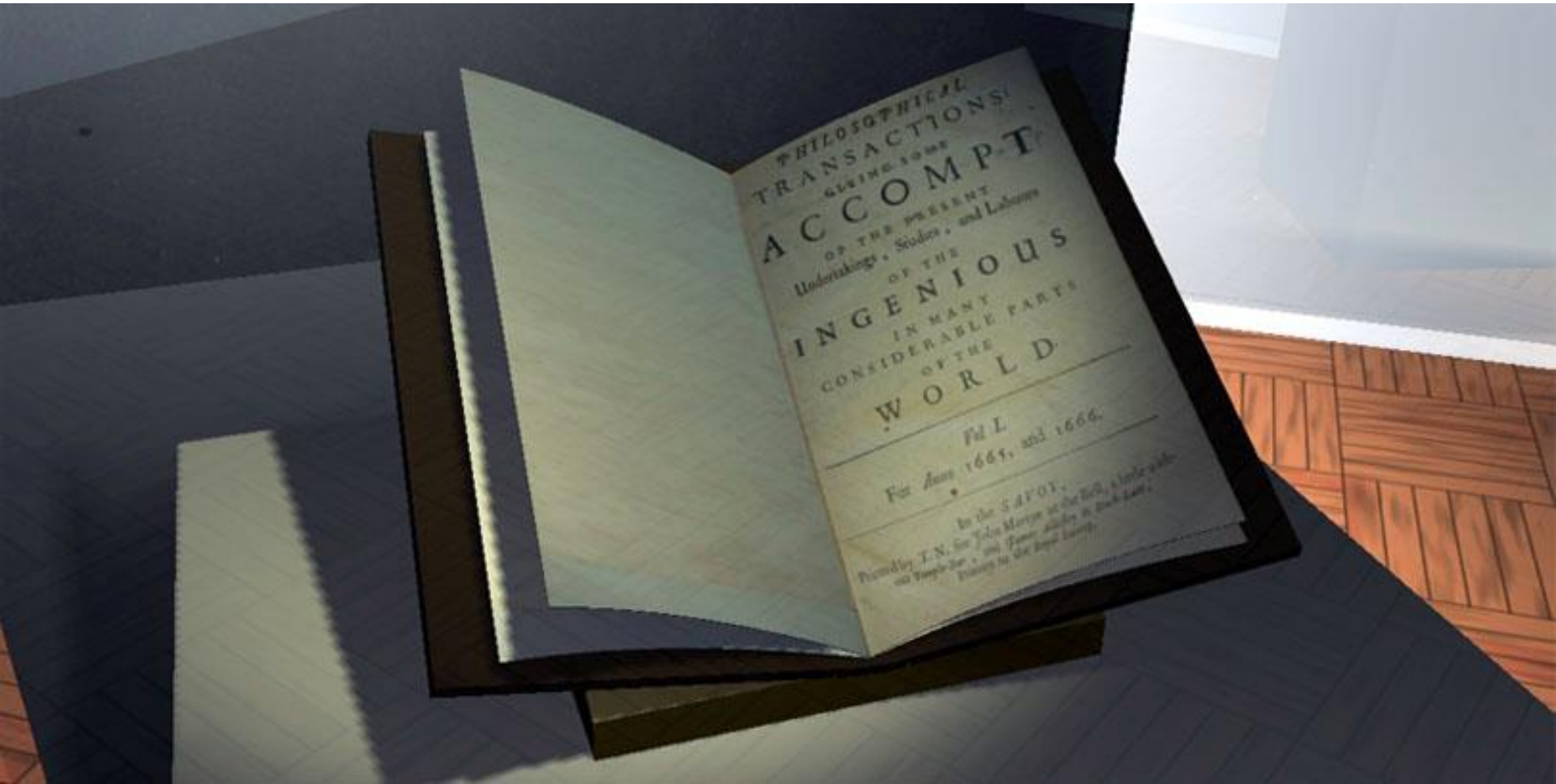
Sharing

Sharing policies are used to manage distribution of data between collaborators and time-dependent access controls.

Federation

Federation policies are used to manage interactions between independent data management systems, including authentication, synchronization of holdings, replication of data, and sharing of data.

What the philosopher might think...



Central to the whole manifesto of Wilkins' Oxford Club, then of the Royal Society after 1660, was the need to broadcast scientific discovery, useful inventions, and ideas, far and wide. Their declared enemies were the closet scholars of "the schools" who believed true knowledge was only for a learned elite, and could only be got at through classical texts or deductive logic. Men who dealt only in "words" and not in "things", and whose intellectual systems were often circular, obscure and of no benefit to the wider community. The Royal Society, by contrast, dealt in "things": objects, experiments, public testing, and what we now call research.

What the sociologist might think...

1. **'Communalism'** refers to the claim that research results are property of the community
2. **'Universalism'** means that everybody should be able to contribute, for example independent of cultural or national origin
3. **'Disinterestedness'** requires the greater scientific good to be valued higher than personal interests
4. **'Skepticism'** implies critical scrutiny of research results

Merton, Robert K. *The sociology of science: Theoretical and empirical investigations*. University of Chicago press, 1979.

(OPEN) DATA ARE REQUIRED TO ENABLE 'GOOD' RESEARCH

What the scientist might think...

🔄 eprints.soton.ac.uk/257648/1/The_Data_Deluge.pdf

The Data Deluge: An e-Science Perspective

Tony Hey^{*,+} and Anne Trefethen^{*}

UK e-Science Core Programme

Tony.Hey@epsrc.ac.uk

Anne.Trefethen@epsrc.ac.uk

Abstract

This paper previews the imminent flood of scientific data expected from the next generation of experiments, simulations, sensors and satellites. In order to be exploited by search engines and data mining software tools, such experimental data needs to be annotated with relevant metadata giving information as to provenance, content, conditions and so on. The need to automate the process of going from raw data to information to knowledge is briefly discussed. The paper argues the case for creating new

EXAMPLES, CHRONOLOGICALLY

Project Policy: Human Genome Project

1996

NHGRI Policy Regarding Intellectual Property of Human Genomic Sequence

 [Share](#)  [Print](#)

Policy on Availability and Patenting of Human Genomic DNA Sequence Produced by NHGRI Pilot Projects (Funded Under RFA HG-95-005)

April 9, 1996

This document describes the policy of the National Human Genome Research Institute (NHGRI) with respect to availability and patenting of human genomic DNA sequence produced under grants funded as a result of RFA HG-95-005. In conformity with the existing spirit and philosophy of the Human Genome Project (HGP) and in response to the recommendations of advisors and the expressed wishes of the community, NHGRI seeks to make DNA sequence information **rapidly and freely** as possible.

Background

The HGP is an international research effort, begun in 1990, which has the scientific goals of generating maps of the human genome [1] and producing the complete sequence of the human DNA by the year 2005. The project was undertaken in the U.S. following the advice of several scientific committees that emphasized its importance in creating a resource that "will facilitate research in biochemistry, physiology and medicine"[2], "have a major impact on health care and disease prevention" [2] and provide "enormous scientific and technological advances ... having both basic and commercial applications" [3]. At the National Institutes of Health (NIH), the National Human Genome Research Institute (NHGRI) was founded to implement the HGP.

The HGP has progressed rapidly, even beyond optimistic expectations. The initial mapping goals are nearly completed and recent improvements in DNA sequencing technology and capacity have led many scientists, including NHGRI advisors, to conclude that complete sequencing of human genomic DNA should begin. Early in 1995, NHGRI issued RFA HG-95-005 to

Funder Policy: DFG ¹⁹⁹⁸

Recommendation 7: Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.

DFG

Recommendations of the Commission on Professional Self Regulation in Science

**Proposals for Safeguarding Good Scientific Practice
January, 1998**

Project policy: Tools 2001

genomics.energy.gov

Human Genome Project Information • Genomic Science Program • DOE Microbial Genomics • home



Human Genome Project Information

About the HGP	Ethical / Legal Issues		Medicine	Education	Gene Gateway	Research Archive
Goals	History	Timeline	Benefits	ELSI	Genetics 101	FAQs

Announcements on the First Analysis of Genome Sequence

February 12, 2001

DATA RELEASE AND ACCESS PRINCIPLES AND POLICY

- The human genome, the common heritage of all humanity, is arguably the most valuable dataset the biomedical research community has ever known. It holds long-sought secrets of human development, physiology, and medicine.
- The highest priority of the International Human Genome Sequencing Consortium is to make the sequence of the human genome available to the world's scientists rapidly, freely, and with the minimum of restriction.

Since the sequencing phase of the Human Genome Project began, the Consortium has made the sequence of the human genome available to the world's scientists rapidly, freely, and with the minimum of restriction.

<http://www.ncbi.nlm.nih.gov/Genbank/>
<http://www.ebi.ac.uk/embl/index.html>
<http://www.ddbj.nig.ac.jp/>

- <http://www.ncbi.nlm.nih.gov/Genbank/>
- <http://www.ebi.ac.uk/embl/index.html>
- <http://www.ddbj.nig.ac.jp/>

- Translating the text of the human genome into practical applications that will alleviate suffering is one of the greatest challenges facing humankind. This mission will require the work of tens of thousands of scientists throughout the world. No scientist wanting to advance this cause should be denied the opportunity to do so for lack of access to the human genome sequence.

Funder Policy: US NIH 2002

NIH ANNOUNCES DRAFT STATEMENT ON SHARING RESEARCH DATA

Release Date: March 1, 2002

NOTICE: NOT-OD-02-035 (See [NOT-OD-03-032](#) for Update)

National Institutes of Health

Data sharing

of the National Institutes of Health's particularly important for unique data that cannot be readily replicated. Data sharing allows scientists to expedite the translation of research results into knowledge, products, and procedures to improve human health. THE NIH IS DEVELOPING A STATEMENT ON DATA SHARING THAT EXPECTS AND SUPPORTS THE TIMELY RELEASE AND SHARING OF FINAL RESEARCH DATA FROM NIH-SUPPORTED STUDIES FOR USE BY OTHER RESEARCHERS. INVESTIGATORS SUBMITTING AN NIH APPLICATION WILL BE REQUIRED TO INCLUDE A PLAN FOR DATA SHARING OR TO STATE WHY DATA SHARING IS NOT POSSIBLE. This statement will apply to extramural scientists seeking grants, cooperative agreements, and contracts as well as intramural investigators.

Journals, e.g. NAR 2003 (?)

OXFORD JOURNALS

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS

Oxford Journals > Life Sciences > Nucleic Acids Research >

GENERAL POLICIES OF THE JOURNAL

New – Please note that the journal now encourages authors to publish online.

Open Access charges are now payable using our Author Gateway online with a credit/debit card, or request an invoice.

Authors' responsibilities

Submission of a manuscript implies that it reports unpublished work elsewhere. **All files, including figures, must be original** and submitted by the corresponding author. The corresponding author checks on behalf of all authors to indicate that they are in complete agreement and are prepared to abide by the general policies of NAR as published in this journal. The senior author, then full contact details of the latter must be provided in all circumstances the journal may need to communicate with authors. The addresses of all co-authors must also be provided.

During submission, authors are requested to upload a cover page to aid the processing of the manuscript. This **must** include the title, authors, and NAR, either partial or in entirety, which has been rejected. The manuscript number of the earlier submission must be provided to any editorial or referee reports and a summary of the changes. The cover should also contain details of any data obtained from other sources or personal communication(s). The corresponding author must provide details of each inclusion.

Authors must also advise the journal of any related manuscript published in another journal, especially where the related manuscript describes results or interpretation of the current NAR submission. A cover page must be included in the online submission system.

Conflicts of Interest

NAR policy requires that EACH author of all manuscripts reveal any financial interests or connections, direct or indirect, or other situations *that might raise the question of bias* in the work reported or the conclusions, implications, or opinions stated - including pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organization(s), personal relationships, or direct academic

- **Software:** If the manuscript describes new software tools or the implementation of novel algorithms the software must be freely available to users at the time of submission (either as executable versions for multiple, common platforms or as source code). Availability must be clearly stated in the article. Authors must ensure that the software is available for a full TWO YEARS following publication, preferably through a download link on a stable URL. Authors are encouraged to make their source code available through an open source license (see www.opensource.org for examples).
- **Databases:** must be freely available to all via the web and not be password protected. Authors are encouraged to make the contents of their databases freely available as flat or relational files upon request.
- **Web servers:** must be freely available to all with no login requirement. Authors are encouraged to make the underlying code and data for their web tools available upon request.

The Editors are prepared to deny further publication rights in the Journal to authors unwilling to abide by these principles.

Deposition of sequence and structural data

Sequence information, co-ordinates used to create molecular models described in a manuscript, and structural data must be submitted in electronic form, prior to acceptance, to the appropriate database for release no later than the date of publication of the corresponding article in the Journal. Deposition numbers and/or accession numbers provided by the database should be included in the manuscript and entered into the relevant boxes during online submission or communicated to the Executive Editor handling the manuscript as soon as received. In cases where there may be no appropriate database, authors must make their data available on request. Atomic co-ordinates may be included in the publication as supplementary material. Manuscripts will not be published until the Journal is in receipt of the deposition number.

For papers reporting novel nucleic acid sequences

Nucleic acid sequence information must be deposited with one of the three major collaborative databases (EMBL/GenBank/DBJ). For sequences obtained from a public or private web site, it is the author's responsibility to ensure that any sequence used within the manuscript is deposited before publication. It is

Funder Policy: CRCs 2010ff

DFG form 50.06 – 05/12

Page 1 of 17

Guidelines

Collaborative Research Centres

III.1.3 Information Infrastructure Project

The main purpose of an Information Infrastructure Project in a Collaborative Research

Information Infrastructure Project

you may also entail the use and testing or development of virtual forms of scientific communication related to such data. Funding may be available for the development and implementation of a design and for the provision of an efficient information infrastructure to make it feasible.

This is intended primarily to achieve two objectives. On the one hand, it serves to fulfil the requirement, according to the principles of good scientific practice, to archive research data for at least ten years. On the other hand, it is assumed that scientific synergies will be generated in the Collaborative Research Centre through shared data platforms and communication forums as well as the resulting efficient reuse of data.

Professional management of the data collected, processed and archived in the Collaborative Research Centre is expected. As a rule, the Collaborative Research Centre should therefore take advantage of the relevant local information facilities (such as the host university's libraries and computer centres). The reuse of existing tools and technologies should always be given preference over the development of new instruments. Data that have been collected and processed must be archived for the long term to keep them accessible beyond the funding duration.

In detail, an Information Infrastructure Project may focus on one or more of the following objectives:

- *Research data:* Development of a database to store research data generated by the Collaborative Research Centre, including the assignment of metadata. In-

Funder Policy: EPSRC 2011

EPSRC

Engineering and Physical Sciences
Research Council

[Home](#) > [About us](#) > [Our service standards](#) > [EPSRC Policy Framework on Research Data](#) > [Impact, Timescales and Support](#)

IMPACT, TIMESCALES AND SUPPORT




EPSRC recognises that the infrastructure and resources allocated to managing and sharing research data vary widely among the research organisations it funds. EPSRC also acknowledges that to comply with this framework research organisations will need to review their research data management practices and capabilities: change may be required in a number of areas ranging for example from simply clarifying and formalising existing processes and policies to the creation of new posts and investment in new infrastructure. In working towards an optimal approach, it is further anticipated that not all the identified actions/investments may be achievable in the medium to long term.

a clear roadmap

To allow time for this EPSRC expects all those it funds to have their policies and processes with EPSRC's expectations by **1st May 2012**, and to be fully compliant with these expectations by **1st May 2015**. EPSRC will monitor progress and compliance on a case by case basis. EPSRC will investigate non-compliance; if it appears that proper sharing of research data is being obstructed EPSRC reserves the right to impose appropriate sanctions. Support and advice are available to assist organisations working to comply with this framework. In particular EPSRC recommends that further advice be sought from;

JISC (<http://www.jisc.ac.uk/whatwedo/activitiesbytopic>)
the Digital Curation Centre (<http://www.dcc.ac.uk/>)
and/or the Information Commissioner's Office (<http://www.ico.gov.uk/>).

Funder Policy: many... 2012

 Full Coverage
  Partial Coverage
  No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

Institutional Policy: e.g. Edinburgh 2011



The University of Edinburgh

Policies and Regulations

Research Data Management Policy

This policy for managing research data was approved by the University Court on 16 May, 2011.

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
10. Except where the data are of future historical interest, research data should not be handed over to commercial publishers or other third parties unless this is a condition of funding. **rights to reuse** should not be handed over to commercial publishers or other third parties unless this is a condition of funding. make the data openly available for re-use,

Funder Policy: EC 2012



PRESS RELEASE

Brussels, 17 July 2012

Scientific data: open access to research results will boost Europe's innovation capacity

The European Commission today outlined measures to improve access to scientific information produced in Europe. Broader and more rapid access to scientific information will, the Commission will, "improve the efficiency of research and innovation, and thus contribute to the growth and innovation of the European Union."

- define open access to peer-reviewed publications as the general principle in Horizon 2020, either through open access publishing ('Gold' open access) or self-archiving ('Green' open access)
- promote open access to research data (experimental results, observations and computer-generated information etc.) and set a pilot framework in Horizon 2020, taking into account legitimate concerns in relation to privacy, commercial interests and questions related to large data volumes
- develop and support e-infrastructures to host and share scientific information (publications and data) which are interoperable on European and global level


open access to research data

...on that public bodies across the European Union ...an essential element in realising the European ... Research Area (ERA). The Communication and Recommendation on scientific information complement a Communication on "A reinforced European Research Area partnership for excellence and growth", also being adopted today, which sets out key priorities for completing the European Research Area, including the optimal circulation, access to and transfer of scientific knowledge.

The European Commission will continue to fund projects related to open access. In 2012-2013, the Commission will spend €45 million on data infrastructures and research on digital preservation. Funding will continue under the Horizon 2020 programme. During the same period, the Commission will support experimentation with new ways of handling scientific information (e.g. new peer-review methods and ways of measuring article impact).

See also [MEMO/12/565](#)

Government Policy: e.g. UK ²⁰¹²



Inside Government
How government works Get involved Departments Worldwide Topics
Announcements Consultations **Policies** Publications Statistics

Policy

Improving the transparency and accountability of government and its services

Organisations: [Cabinet Office + 1 other](#)
Updated: [2 May 2013](#)
Topic: [Government efficiency, transparency and accountability](#)
Ministers: [The Rt Hon Francis Maude MP](#)
Advisory groups: [Public Sector Transparency Board + 2 others](#)

Policy

Detail

Latest

Commitment to publishing data

[The Open Data Institute](#)

[Releasing data in open formats](#)

[Anonymised formats](#)

Supporting detail:

Commitment to publishing data

The government is committed to publishing public data (between 2010 and 2015) under a **Open Data** statutory publication scheme [scheme](#).

These strategies, along with the [open data white paper](#) commit us to a publishing all public sector data, unless it is private data collected from people or restricted for national security reasons.

Local authorities and arms-length bodies also comply with the [local](#)

International Policy, e.g. G8 ²⁰¹³



- Share information on national research infrastructure priorities and prioritization processes; identify areas of potential benefit that could be achieved through sharing of best practices.
- Create a representative list of GRIs open to global cooperation of interest to new partners.

We invite the GSO to report in 2015 on their progress.

3. Open Scientific Research Data

Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and its results communicated. It can provide society with the necessary information to solve global challenges. We are committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of scientific research are as widely available as practical, enable transparency in science and engage the public. We will develop a set of principles for open science, and further discussions.

- To the greatest extent possible, ensure that publicly funded research data is made available to the public, whilst acknowledging the legitimate concerns of private partners.

publicly funded

SUMMARY AND CONCLUSION

On a data journey

- ~350 years of “research” and ~15 years of research data policy
- Address research conduct, project, technical, funder, institutional, national, international
- Even massive programmes need years to develop practical advise
- Two step process from aspiration to implementation

Conclusion: a two step process

1. Aspirational Policies > necessary “Zeitgeist”
 - For setting the scene
2. Practical Policies > implementing “Seachange”
 - Are close to the research process, either methodologically or locally
 - Include actual infrastructure
 - Include practical advice
 - Include funding