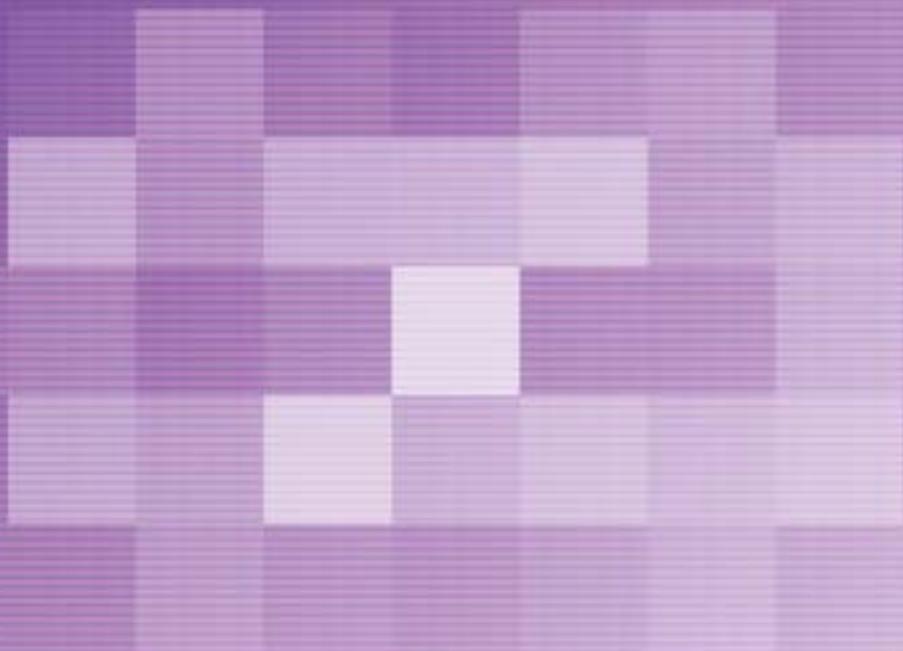


National Statistics Code of Practice

Protocol on Data Management,
Documentation and Preservation



National Statistics Code of Practice

Protocol on Data Management, Documentation and Preservation

© Crown copyright 2004

Published with the permission of the Controller of Her Majesty's Stationery Office (HMSO).

ISBN 0 11 7033820

Applications for reproduction should be submitted to HMSO under HMSO's Class Licence:

www.clickanduse.hmso.gov.uk

Alternatively applications can be made in writing to:

HMSO
Licensing Division
St. Clement's House
2-16 Colegate
Norwich, NR3 1BQ

Contact points

For enquiries about this publication, contact the National Statistics Policy Helpline

Tel: **020 7533 6210**

(for advice on interpreting the Code of Practice)

E-mail: **james.denman@ons.gov.uk**

For general enquiries, contact the National Statistics Customer Contact Centre on:

0845 601 3034

(minicom: 01633 812399)

E-mail: **info@statistics.gov.uk**

Fax: 01633 652747

Letters: Room D115,
Government Buildings,
Cardiff Road,
Newport, NP10 8XG

You can also find National Statistics on the Internet at **www.statistics.gov.uk**

A National Statistics Publication

National Statistics are produced to high professional standards set out in the *National Statistics Code of Practice*. They undergo regular quality assurance reviews to ensure that they meet customer needs.

They are produced free from any political influence.

Navigate through this document by using Bookmarks, Thumbnails or Links from the Contents listing below. There are more sub-categories in the Bookmarks than there are in the Contents listing. Prevent the printing of these instructions by unchecking 'Annotations' in the print dialogue box.

Contents

Summary of Principles	4
Introduction	5
Purpose	5
Compliance	6
A guide to putting the principles into practice	7
Data Management Policy	7
Data Managers – Management of the complete life-cycle	8
Documentation/metadata	9
Integrity and security of statistical resources	12
Retention, preservation and destruction	13
References	16
References	17
Legislation	17
Archiving	18
Best Practice	20
Metadata Standards	21

Summary of Principles

Code of Practice – Statement of Principles

The provision of National Statistics will be effectively managed.

Processes and methods used to produce National Statistics will be fully documented.

Statistical data will be treated as valuable and irreplaceable, with their value increasing through widespread and long-term use.

National Statistics will archive material likely to be of historical interest, subject to security, confidentiality and statutory obligations.

Additional principles embodied in this Protocol

Each organisation responsible for producing National Statistics will have a Data Management Policy with the aim of ensuring the most effective use of its statistical resources by present and future generations.

Responsibility for managing the complete life cycle of each statistical resource will be vested in one or more identifiable data managers.

Statistical resources will be documented in a standard manner. Documentation will include easily accessible metadata designed to foster greater awareness, usability and understanding of the data and enhance their functionality.

All organisations responsible for producing National Statistics will protect the integrity and security of their statistical resources, and minimise the risk of loss or disclosure.

Data managers will archive material likely to be of historical interest or essential to the analysis of time-series – subject to considerations of security, confidentiality, statutory obligations and costs.

Introduction

For background information on this document please see *An Introduction to the National Statistics Code of Practice and its supporting Protocols*.

Purpose

This protocol sets out how the producers of National Statistics will carry out their responsibilities for managing, documenting, retaining and preserving the statistical resources which they control – as outlined in the *Framework for National Statistics* and the *National Statistics Code of Practice*.

Government statisticians manage a huge array of irreplaceable statistical resources. This bank of knowledge is fundamental to the strategic aims of informing the nation, and supporting decision-making. It can also be shared with, and enhanced by, other organisations to provide benefits both now and in the future.

If this resource is to be used in the most effective and efficient way it must be managed in a clear and accountable manner. With proper management and documentation procedures in place, data suppliers will have an audit trail to assure themselves that the data they have provided is handled in a way which accords with their agreed condition of use, without risk to confidentiality, copyright, or intellectual property rights, and in compliance with all statutory and non-statutory obligations. Robust procedures will also reassure stakeholders that the data will be used and re-used in an optimum manner, and for the benefit of users and everyone in society, whether now or in the future.

Compliance

The best practice principles set out in the Protocol may require producers in government departments and agencies to develop and establish new systems and new procedures. Compliance may, therefore, be an incremental process dependent on cost constraints and competing priorities. Furthermore, it may not be possible for producers to apply these principles fully to all the systems from which statistics are derived – a qualification which applies, in particular, to management or administrative systems.

A guide to putting the principles into practice

1. Each organisation responsible for producing National Statistics will have a Data Management Policy with the aim of ensuring the most effective use of its statistical resources by present and future generations.

National Statistics constitute valuable and irreplaceable assets and should be managed in a way that encourages their widest possible use and re-use. This will be achieved through effective data management procedures designed to ensure that statistical resources not only meet current needs but can also be exploited in the future.

- a) Each organisation responsible for National Statistics will adopt a Data Management Policy designed to ensure that the management of their statistical resources meets the standards set out in this Protocol.
- b) Each organisation's Data Management Policy should embrace the whole life-cycle from data acquisition through to long-term preservation or possible destruction, and should aim to:
 - foster the development, implementation and maintenance of the detailed data policies, standards, and procedures set out in the *National Statistics Code of Practice* and all of its supporting Protocols;
 - align the organisation's data management arrangements with wider Government policies, and with its statutory obligations – see 'References';

- ensure that all staff receive appropriate guidance and training to enable them to fulfil their data management obligations;
 - provide the information required to maintain a central repository of metadata covering all National Statistics;
 - incorporate guidance on the retention and preservation of both electronic and paper records.
- c) The application of this policy to administrative sources of data will recognise the administrative uses of those data as well as the statistical purposes.
2. **Responsibility for managing the complete life-cycle of each statistical resource will be vested in one or more identifiable data managers.**
- a) The National Statistician, Heads of Profession, or Chief Statisticians in the devolved administrations, will vest responsibility for the management of each statistical resource in one or more publicly identifiable data managers.
 - b) Nominated managers will be responsible for the following range of tasks:
 - i. Security Control – compliance with corporate arrangements for ensuring that statistical resources are not put at risk or corrupted at any stage in their life-cycle and meet relevant security and confidentiality restrictions.
 - ii. Statutory Compliance – compliance with the organisation’s statutory and regulatory requirements.
 - iii. Code Compliance – compliance with the principles and standards set out in the *National Statistics Code of Practice* and its supporting Protocols.

- iv. Documentation – provision of full documentation on the major stages in the life-cycle of each resource including easily accessible and comprehensive metadata, using standard templates where appropriate, in order to foster users' awareness and understanding.
 - v. Access – managing appropriate access to data in accordance with the guarantees given to data contributors.
 - vi. Retention, Preservation and Archiving – compliance with corporate retention and preservation policies in order to ensure the continuing, and long-term, functionality and availability of both the resources themselves as well as any contextual and allied material.
- c) Where a statistical resource contains personal information and is, therefore, required to be registered under the Data Protection Act, the responsible data manager will operate in accordance with the procedures set out by their organisation's Data Controller(s) in order to ensure that legal requirements are met.
3. **Statistical resources will be documented in a standard manner. Documentation will include easily accessible metadata designed to foster greater awareness, usability and understanding of the data and enhance their functionality.**

It is important to maintain full and comprehensive documentation in order, for example, to:

- safeguard knowledge and expertise associated with each resource;
- fulfil statutory requirements relating to confidentiality and access;
- provide management information about each resource;
- facilitate the integration and cross-analysis of different datasets;

- provide user-friendly information about each resource;
 - enable the long-term preservation, retention and subsequent retrieval of each resource.
- a) Data managers will document their statistical resources using a standard template. This documentation will include details of the arrangements they have in place to cover, for example:
- i. data collection (including a business justification);
 - ii. consultation with respondents and users;
 - iii. quality and methodology;
 - iv. security, confidentiality and disclosure;
 - v. dissemination;
 - vi. archiving – including the identification of resources selected as permanent records under the relevant Public Records legislation – see ‘References’.

Consistency

- b) Wherever practicable, documentation will conform to the National Statistics standard and will be compiled in accordance with central government guidelines.
- c) Documentation will be validated and quality checked on a regular basis to avoid the degradation of detail over time. Contact details should be reviewed on a regular basis to reflect staff changes.

Review

- d) Documentation requirements for both live and preserved data will be reviewed regularly in the light of emerging standards in order to ensure that they continue to meet producers’ and users’ needs.

Metadata

Metadata – or ‘data about data’ – is a specific form of documentation and refers to the information that is made available to users in order to improve their understanding of the data. Comprehensive and complete metadata helps users to make informed and full use of data holdings and minimises the likelihood of misuse.

- a) Every organisation’s Data Management Policy will set out the arrangements it has in place for compiling and maintaining metadata covering all of its resources.
- b) Metadata will be compiled using standard templates and terminology and in accordance with central government guidelines and common standards and definitions, and will be closely linked to the data described.
- c) Metadata compilation on all new National Statistics will be completed in time for their first release.
- d) Metadata will be made available to the widest possible audience through appropriate electronic portals and other dissemination channels.

Discovery metadata

- e) ‘Discovery metadata’ – information enabling users to identify and access information relevant to their needs – will be compiled for every resource and will be made available in a comprehensive catalogue – in both paper and electronic format. The information will be kept up to date, be easily accessible and will contain information about the title, content, geographic context, timeliness, availability and accessibility of each resource together with appropriate ‘Contact’ details. Access will be enhanced through the provision of indexes, and availability through web-based search engines.

Statistical metadata

- f) Additional metadata will be made available to users concerning the more technical details of each resource. This will comprise, among other things, a description of the data collection arrangements including sample design, questionnaires, coding instructions and classifications, editing, validation, methodologies, reason for and method of data collection, quality of data, confidentiality and anonymisation procedures and any other relevant materials.

Technical metadata

- g) Metadata attached to electronically held data which is to be preserved permanently, or for a significant length of time, will include information about the software used to arrange and process the resource. This is to ensure that the data can continue to be understood, manipulated and accessed over time.
4. All organisations responsible for producing National Statistics will protect the integrity and security of their statistical resources, and minimise the risk of loss or disclosure.

Data Integrity and Security

- a) Every organisation's Data Management Policy will set out the arrangements it has in place for ensuring the integrity and security of its data holdings in order to preserve the guarantee of confidentiality given to contributors, and ensure compliance with the Data Protection Act.

Physical security of information holdings

- b) Statistical resources can be damaged or lost because of human error, because technical problems lead, for example, to the corruption of data files, or because of disasters such as fire or flood. All those with data management responsibilities will follow their organisation's standard procedures for ensuring the physical security of their resources along with associated back-up arrangements for minimising the impact of such events. These arrangements should reflect the importance of each resource and the impact of its destruction.

- c) Where National Statistics activities are outsourced, these requirements will be set out in contractual documentation.
 - d) The National Statistician should be notified of the loss of, or damage to, any resource that for any reason is likely to affect significantly financial markets or public policy.
 - e) The security of every live statistical resource will be underpinned by back-up procedures.
 - f) Preserved data should be periodically restored to test their integrity.
5. **Data managers will archive material likely to be of historical interest, or essential to the analysis of time-series – subject to considerations of security, confidentiality, statutory obligations and costs.**
- a) National Statistics constitute valuable and irreplaceable assets whose value can increase through widespread and long-term use. Each organisation's Data Management Policy will set out the arrangements it has in place for the retention, long-term preservation, and destruction of its resources.
 - b) Each Head of Profession will formulate and operate a corporate selection policy, involving users and experts as appropriate, and in consultation with the appropriate national archive, in order to determine which of their resources should be preserved. Their selection policy should satisfy the requirements of the Public Records Acts and they should choose a place of deposit recognised by their appropriate national archive – see 'References'.
 - c) Each organisation's retention and preservation policy should take into account the relative value and rarity of the material, the amount and type of use it gets, the way in which it has been stored and its physical condition. Decisions on retention should pay due regard to:

- i. The potential use of the data in supporting future research, taking account of the value to be derived from retaining records of individual responses over and above published aggregates, and the ability to derive new time series;
- ii. the anticipated historical value of the information resource – for example its relevance to contemporary events or its ability to mark turning points in the social and economic life of the nation;
- iii. considerations of storage, maintenance, and protection from damage;
- iv. considerations of security, confidentiality, statutory obligations and cost constraints including the additional costs of converting paper documents for electronic storage and archiving;
- v. the fact that some statistical products exist only on the Web.

Preservation and retention

- d) The selection policy should cover both the full version of the data as well as any anonymised datasets where applicable.
- e) References to printed publications and their source will be kept with the digitised information.
- f) All formal printed publications will be preserved as a permanent record within one month of publication by depositing them with the relevant legal deposit office. See ‘References’.
- g) Records worthy of permanent preservation including datasets and relevant documentation will be deposited with the relevant national archive. See ‘References’.

- h) Other data deemed worthy of retention for a reasonable amount of time, whether microdata or aggregate data, should also be preserved and made available to users where confidentiality constraints allow. This may be through the organisation's own website or through a suitable archiving service. See 'References'.
- i) Electronic records retained by National Statistics organisations themselves should be held in a form accessible by contemporary software that is not application-dependent. Significant changes in either hardware or software should be accompanied by comprehensive testing of the accessibility and integrity of preserved data in the new environment before the old environment is irrecoverably lost.
- j) A resource should be retained if it is the only known source of statistics on a particular topic depending on the degree to which it is complete or adequately documented.

Destruction

- k) Decisions on destruction will accord with procedures approved by the appropriate national archive and will be taken in the context of each organisation's policy on retention and preservation.
- l) Data not selected for permanent preservation should be destroyed when of no further use and in accordance with the guidance produced by the appropriate national archive. See 'References'.
- m) Any decision concerning the destruction of a particular statistical resource should be taken in the context of the *Protocol on User Consultation*.
- n) Data managers will act to ensure that any destruction of a statistical resource is complete by safely destroying all copies of the dataset held by the organisation, including back-up files.

References

The Principles of Good Data Management – Intra-governmental Group on Geographic Information (IGGI) – September 2000
www.iggi.gov.uk/achievements_deliverables/pg_dmg.htm

The Principles of Good Metadata Management – IGGI – January 2002
www.iggi.gov.uk/achievements_deliverables/prinmeta.htm

Preserving and Sharing Statistical Material – The Royal Statistical Society and The UK Data Archive – 2002 – ISBN 0-906805-02-3
www.data-archive.ac.uk/home/PreservingSharing.pdf

References

Where there is a reference in this Protocol to the ‘appropriate national archive’, then for UK Departmental datasets that are primarily about the geographical area of a devolved administration, the appropriate national archive may include the relevant devolved archive.

Relevant Legislation:

Public Records Acts 1958 and 1967
For Northern Ireland – the Public Records Act (NI) 1923
For Scotland – Public Records (Scotland) Act of 1937.
Health and Safety at Work Act 1974 (and related regulations)
Unfair Contract Terms Act 1977
Limitation Act 1980
Companies Acts 1985 and 1989
Financial Services Act 1986
Latent Damage Act 1986
Consumer Protection Act 1987
Copyright, Designs and Patents Act 1988
Value Added Tax Act 1994
Civil Evidence Act 1995
Data Protection Act 1998
Human Rights Act 1998
Freedom of Information Act 2000
Freedom of Information (Scotland) Act 2002
Copyright Acts
Modernising Government Framework (1999)

Copies of Acts from 1988 onwards can be obtained through the HMSO website:
www.legislation.hmso.gov.uk

Eurostat legislation can be found on the Eurostat website:
<http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat>

Archiving

The National Archive (TNA):
(Formerly the Public Records Office (PRO))
National Archives (Public Records Office)
Kew, Richmond
Surrey, UK. TW9 4DU
Telephone: 020 8876 3444
Fax: 020 8392 5286
Minicom: 020 8392 9198
enquiry@nationalarchives.gov.uk

National Archives (formally the Public Records Office)
PRO website
www.pro.gov.uk

Contact Details for Departments and Projects at the National Archives
www.pro.gov.uk/contacts/contacts.htm#rmd

Digital Archive
For enquiries concerning the transfer of Government electronic records:
Public Record Office
Kew, Richmond
Surrey UK. TW9 4DU
digital-archive@nationalarchives.gov.uk

National Archives Standards and Guidance
www.pro.gov.uk/recordsmanagement/standards/default.htm

Standards for Records Management – Standards for the management of Government records

The Public Record Office produces standards and best practice guidance on all aspects of records management, both paper-based and electronic:

www.pro.gov.uk/recordsmanagement/standards/introduction.htm

www.pro.gov.uk/recordsmanagement/standards/default.htm

www.pro.gov.uk/recordsmanagement/standards/documentation.pdf

International Standard ISO15489

The International Organisation for Standards' work on Records Management:

www.iso.ch/iso/en/ISOOnline.frontpage

National Archives of Scotland (NAS):

Website for NAS

www.nas.gov.uk/

NAS Conservation/Preservation

www.nas.gov.uk/conservation.htm

The Scottish Records Advisory Council

www.nas.gov.uk/about_nas.htm

National Library of Scotland (NLS) for Scottish publications

www.nls.uk/

Public Records Office Northern Ireland (PRONI):

Northern Ireland Records Management Standard (NIRMS)

<http://proni.nics.gov.uk/nirms/edrm.htm>

Public Records Office Northern Ireland (PRONI)

<http://proni.nics.gov.uk/>

PRONI Records Management

<http://proni.nics.gov.uk/NIRMS/edrm.htm>

Other Archiving Organisations:

UK Data Archive (Essex Data Library)

www.data-archive.ac.uk/

Edinburgh University Data Library

<http://datalib.ed.ac.uk/>

Best Practice**IS/IT management**

The Office of the e-Envoy publishes IS/IT-related Framework Policies which apply to the whole of central government and the rest of the public sector. Policies published to date include:

Electronic Records Management

Metadata

Authentication

Security

The Office of the e-Envoy also publishes Guidance and Guidelines on IS/IT-related topics. These currently include:

Guidelines for UK government websites

e-Government metadata standard

The policy and guidance material can be accessed from:

www.e-envoy.gov.uk/Home/Homepage/fs/en

Common metadata standards**Dublin Core**

<http://dublincore.org/>

Data Documentation Initiatives (DDI)

www.icpsr.umich.edu/org/init.html

www.icpsr.umich.edu/DDI/index.html

e-Government Metadata Standard (e-GMS)

www.govtalk.gov.uk/schemasstandards/metadata.asp

The GCL (Government Category List)

www.govtalk.gov.uk/schemasstandards/gcl.asp

