IGGI

Intra-governmental Group
on Geographic Information

The Principles of Good Data Management

The Principles of Good Data Management

2nd Edition

July 2005

Office of the Deputy Prime Minister: London

The Office of the Deputy Prime Minister
Eland House
Bressenden Place
London SW1E 5DU
Telephone 020 7944 4400
Web site www.odpm.gov.uk

This publication is also available at www.iggi.gov.uk

Any enquiries relating to this publication should be addressed by email to:
iggi@odpm.gsi.gov.uk

Further copies of this publication are available from:
ODPM Publications
PO Box 236
Wetherby
West Yorkshire
LS23 7NB
Tel: 0870 1226 236
Fax: 0870 1226 237
Textphone: 0870 1207 405
E-mail: iggi@odpm.gsi.gov.uk

Printed in Great Britain on material comprising 80% post-consumer waste and 20% ECF pulp.

July 2005
Product Code: 05 PLUS 03146

Acknowledgements
This guide, version 2.0, dated June 2005, was prepared by the IGGI Working Group on Data Management and Standards. Thanks to Jeremy Giles, British Geological Survey and his cross-departmental team for their considerable efforts in putting together this guide. Thanks also to David Lowe, British Geological Survey, for editing the document.

FOREWORD

By Philip White, IGGI Chairman

The Good Data Management booklet was originally produced 2002 as part of IGGI's commitment to encourage best practice when dealing with government geographic information. Recognising that the environment is constantly evolving, the guide has been revised to take account of current best practice, recent developments and reflect key legislation. In recent years, there has been an increased emphasis on facilitating data sharing both within and between organisations. The principle of "collect once, use many times" is well established as a concept but can only be achieved with data management. This booklet complements The Principles of Metadata Management also published by IGGI.

The booklet provides best practice guidance for those responsible for managing data of a geographic nature. Specifically, it considers the benefits, drivers, principles and mechanisms needed for good data management.

This guide helps ensure that processes are in place for initial input and maintenance of reliable metadata for retrieval through ODPM's Maps on Tap and the gigateway Data Locator. If the metadata policies recommended in this leaflet are followed, users can expect consistency in the data they are using and exchanging with their colleagues throughout central government.

I would like to thank the members of the group chaired by Jeremy Giles from British Geological Survey for their hard work and the commitment involved in completing this guide.

CONTENTS

1 Purpose of this guide

This guide provides general guidance on the management of data. The guide has been produced for those responsible for geographic information, although the principles are equally relevant to other types of government information.

Government departments and agencies collect, generate, store and use large amounts of data that have been obtained at considerable cost. Much of this data is geographical in that they are referenced to geographical locations, such as points, lines or areas; e.g. post codes.

The importance of good Data Management has become increasingly recognized over recent years and a body of legislation reflects this change in attitude. Key elements of the relevant legislation include:

- The Freedom of Information Act 2000[1]
- The Environmental Information Regulations[2]
- The Human Rights Act 1998[3]
- The Data Protection Act 1998[4]
- The Public Records Act 1958[5]

The European Commission is also having greater impact, as INSPIRE and the Directive on Public Sector Information demonstrate. These and other drivers mean that Data Management, and the associated activity of records management, need to be given priority by public bodies.

2 What is Data Management?

Data Management is a group of activities relating to the planning, development, implementation
and administration of systems for the acquisition, storage, security, retrieval, dissemination, archiving and disposal of data. Such systems are commonly digital, but the term equally applies to paper-based systems where the term records management is commonly used. The term embraces all forms of data, whether these datasets are simple paper forms, the contents of relational databases, multi-media datasets such as images, or scientific data such as seismic records of the UK land mass.

The management of geographic data is in many ways no different to the management of other types of data. However, it is important to recognise that there may be geography-specific issues that need careful thought as part of Data Management activities; for example, ensuring that any geographic identifiers used are appropriate and resilient. Bearing in mind that one of the strengths of geographic data is the ability to link seemingly disparate pieces of information, it is absolutely critical to ensure that the chosen geographic identifiers allow this.

Key Data Management activities include:

- Data Policy development;
- Data Ownership;
- Metadata Compilation;
- Data Lifecycle Control;
- Data Quality; and
- Data Access and Dissemination.

This guide covers only the key aspects of Data Management. Extensive additional advice is now available from The National Archives[6], The Department for Constitutional Affairs[7] and UKgovtalk[8].

3 Why do we need to manage our data?

Government owns huge amounts of irreplaceable Geographic Information, potentially of use to a wide range of bodies, and there are increasing pressures on departments/agencies to manage these data properly. Examples of these pressures are identified below.

Key drivers for improved Data Management

The Freedom of Information (FoI) Act 2000 came into force in January 2005. General guidance on FoI is provided by The Department for Constitutional Affairs[9] and The National Archives (TNA)[10]. The FoI defines clear duties and responsibilities for all who manage public sector information in all its forms. The Act confers two statutory rights on applicants:

- To be told whether or not the public authority holds that information; and if so,
- To have that information communicated to them.

The Act allows twenty days for a response to be prepared. This limited time means that good Data Management practices must be in place to ensure that public bodies can meet the requirements of the FoI.

The FoI gives rights of access to a wide range of information. However, rights of access to environmental information are provided by a separate statutory regime, the Environmental Information Regulations (EIR). The aim of the regulations, which also came into force in January 2005, is to ensure access for the public to environmental information to enable them to participate in decision-making and obtain justice in environmental matters. This is seen as

essential for creating transparency and building trust within communities and between individuals and public authorities. Guidance on EIRs is provided by TNA[11].

The Human Rights Act 1998 and The Data Protection Act 1998 both provide for the protection of personal information from inappropriate use and the right of access to data held about the individual. These two acts place specific duties on Data Management concerning security and access to personal information.

Other, non-legislative drivers include:

- Increasing recognition that Government data, collected at public expense, must be properly managed in order to realize their full potential and justify their considerable production and maintenance costs.

- Increasing pressure from customers for easier and quicker access to the right information at little or no charge.

- Interoperability between systems and services, for so long seen as desirable, is now becoming a reality. The outputs and credibility of such services depend heavily upon the quality of the data provided. As the number of interoperable services increases, so too does the requirement to have ready access to data of known (maintained) quality.

- Stronger emphasis within Government on the need to rationalize and combine data in order to improve efficiency and add value.

- More reluctance from suppliers to provide data at affordable prices. Stricter control is required by Data Owners over the use of their data to safeguard their Intellectual Property Rights (IPR) and the confidentiality of sensitive data.

4 Benefits of good Data Management

Data Management policies and procedures ensure that data on all media are treated as a valued resource. Implementing such policies and procedures will give many benefits:

Benefits to Data Suppliers

- An increased confidence and trust that their data will be used according to their agreed conditions of use, without risk to confidentiality, copyright or IPR, and in compliance with all statutory and non-statutory obligations.

- Providing a clear understanding of the use of their data, formally documented in a Memorandum of Agreement signed by both supplier and user.

- A fair return for the use of the data they have supplied.

Benefits to Data Brokers/Intermediaries

- Better quality, harmonized and coherent data from the use of common definitions, including geographic references, formats, validation processes and standard procedures.

- Better care of the data holdings through the use of effective data policies and best practice guidance.

- Better control over the data by the clear definition and use of the procedures for the care of data.

- Improved knowledge and understanding of data holdings, their availability, interpretation and use, with subsequent reduction of the risk of duplication or loss, through better cataloguing, metadata and, in time, better access to data via an integrated data environment.

- Improved business processes, including better and more efficient use and re-use of data, and the standardization of datasets that are frequently used by different parts of an organization.

- Increased confidence that the organization complies with statutory and non-statutory obligations, by the regular use of centrally coordinated, frequently updated guidance, codes of practice and training on legal, contractual and other obligations.

- Better control over access to data, both for internal and bona fide external customers, resulting from better data organization and maintenance following defined policies on release, disclosure control and data security.

- More sensible and consistent data charges and conditions of use, resulting from clear pricing and dissemination policies that recognize the need for free access by appropriate customers whilst recovering the appropriate income from customers who seek to make commercial gain.

- An increasing confidence by the customer in the quality of the data managed and in the reliability of outputs that are produced.

Benefits to users and customers

- Improved awareness and understanding of what data are available for current and future use, resulting from better cataloguing and data archiving.

- Improved access to data, free from unnecessary obstacles, safeguarded from disclosure of personal information or infringement of legal and contractual obligations.

- Better quality and more timely information i.e. access to the right information at the right time, resulting from quicker identification of customer needs and the avoidance of wrong or conflicting information, through the use of effective metadata.

- Better value for money, resulting from clear, fair and consistent data charges and conditions of use, which recognize the need for free access by the appropriate customers.

- Better exploitation of data generally, enabled by easier data exchange and integration with other harmonized data.

- Efficiency gains across government and its agencies resulting from the use of better quality data.

5 Principles of Good Data Management

Good Data Management is essential for the effective use of the information resources of public bodies in all their forms. Section 2, above, identified a range of key Data Management activities; these are discussed below.

The key principles of Data Management are illustrated in Figure 1 and described in the text.

Avoid re-collecting data

The largest potential for waste in Data Management is reacquiring an existing dataset. This has been done frequently by public and private sector organizations and must be avoided. In the USA, Executive Order 12906[12] requires government agencies to put internal procedures in place to ensure that they check whether other agencies have already collected information they plan to acquire. Whereas no equivalent instruction exists in the UK, it should be regarded as best practice to use the gigateway[13] Data Locator to search for existing geospatial datasets before new ones are created.

Data lifecycle control

Good Data Management requires that the whole life cycle of datasets be managed carefully. This includes:

- Business justification, to ensure that thought has been given to why new data are required rather than existing data amended or used in new ways, how data can be specified for maximum use including the potential to meet other possible requirements, and why the costs of handling,

storing and maintaining these data are acceptable and recoverable.

- Data specification and modelling, processing, database maintenance and security, to ensure that data will be fit for purpose and held securely in their own databases.

- Ongoing data audit, to monitor the use and continued effectiveness of the data.

- Archiving and final destruction, to ensure that data are archived and maintained effectively until they are no longer needed or are uneconomical to retain.

Figure 1: Key Principles of Data Management

Data policy

The fundamental step for any organization wishing to implement good Data Management procedures is to define a Data Policy. The document may have different names in different public bodies but in each it should be a set of broad, high-level principles that form the guiding framework within which Data Management can operate. This is the document that is approved at senior levels in the public body, and the senior executive who owns the policy (Data Management Champion) manages the resources for its implementation. Section 6 includes a model Data Policy Statement.

Data ownership

One key aspect of good Data Management is the clear identification of the owner of the data. Normally this is the organization or group of organizations that originally commissioned the data acquisition or compilation and retains managerial and financial control of the data. The Data Owner has legal rights over the dataset, the IPR and the Copyright.

Data ownership implies the right to exploit the data, and if continued maintenance becomes unnecessary or uneconomical, the right to destroy them, subject to the provisions of the Public Records and Freedom of Information acts. Ownership can relate to a data item, a dataset or a value-added dataset. IPR can be owned at different levels. For example, a merged or value-added dataset can be owned by one organization, even though other organizations own the constituent data. If the legal ownership is unclear, there are risks that the data can be wrongly exploited, used without payment of royalty to the owner, neglected or lost.

It is therefore important for Data Owners to take action to establish and document:

- The ownership, IPR and Copyright of their data so that these can be safeguarded.
- The statutory and non-statutory obligations relevant to their business to ensure that the data are compliant.
- The departmental policies for data security, disclosure control, release, pricing and dissemination.
- The agreement reached with users and customers on the conditions of use in a signed Memorandum of Agreement, before data are released.

Metadata

All datasets must have appropriate metadata compiled for them. At the simplest level metadata are "data about data". Metadata provide a summary of the characteristics of a dataset. A good metadata record enables the user of a dataset or other information resource to understand the content of what they are reviewing, its potential value and its limitations.

There are many metadata standards, but the ones that are most appropriate to GI are:

- ISO 19115:2003[14] (Geographic Information – Metadata); and
- UK GEMINI – (Geo-spatial Metadata Interoperability Initiative)
  The profile is the result of a collaboration between the AGI[15] and the e-Government Unit[16]. A profile is a subset of

one or several information standards that adopts elements, structures or rules for different user communities. Adherence to the UK GEMINI profile, which will replace the *gigateway Discovery Metadata Specifications* (the NGDF Standard) as the UK's national geospatial metadata profile, allows for the creation of discovery metadata with both ISO 19115 (*Geographic Information – Metadata*) and the national *e-Government Metadata Standard* (eGMS), ensuring compliance with both.

Comprehensive advice on the compilation of metadata can be found in the IGGI booklet entitled "The Principles of Good Metadata Management[17]", the second edition of which was published in May 2004.

Data quality

Good Data Management also ensures that datasets are capable of meeting current needs successfully and are suitable for further exploitation. The ability to integrate data with other datasets is likely to add value, encourage ongoing use of the data and recover the costs of collecting the data. The creation, maintenance and development of quality data require a clear and well-specified management regime.

Data Steward

All datasets need to be managed by a named individual referred to here as the Data Steward; also known as dataset manager and data custodian. A Data Steward should be given formal responsibility for the stewardship of each major dataset. They should be accountable for the management and care of the data holdings assigned to them, in line with the defined data policy. Section 6 provides a list of the responsibilities of the Data Steward.

Data Management Plan

The Data Steward is responsible for the development of a Data Management Plan for each dataset under their responsibility. The objective of the Data Management Plan is to ensure:

- That the dataset is fit for the purpose for which it is required.
- That the long-term management of the dataset is considered for potential re-use.

The individual management plans should be compliant with the local data policy and include:

- Scope of the plan
- Link to metadata
- Responsibilities
- IPR and Copyright
- Quality objectives
- Standards (International, National and local) adopted during compilation of the data
- Staff resources required to manage the dataset
- Physical resources required to manage the dataset
- Long term management of the dataset

Data Management procedures

Individual datasets may require compilation of specific Data Management procedures. These may be needed where specific datasets require detailed operational procedures to ensure their quality; examples of this include scientific and statistical datasets.

Data access and dissemination

Although this aspect will depend upon the business and the financial policy of the organization, the following guidance should be followed.

- Public access to data should be provided in line with The Freedom of Information Act, The Data Protection Act and The Human Rights Act.
- IPR and Copyright of datasets owned by public bodies must be protected, as data should be regarded as an asset.
- IPR and Copyright of third-party data must be respected.
- The potential for commercial re-use and exploitation of the dataset should be considered.
- The right to use or provide access to data can be passed to a third party, subject to agreed pricing and dissemination policies.
- Consideration should be given to the impact of European developments such as the Public Sector Information Directive and INSPIRE.

Data audit

Data Management audits are recommended to ensure that the management environment for given datasets are being maintained. Their purpose is to provide assurance to the Data Management Champion that the resources expended are being used appropriately. Audits of major datasets should be commissioned to ascertain the level of compliance with data policies and the Data Management plans and procedures that have been prepared.

6 Establishing a Data Policy

IGGI has prepared the following model Data Policy Statement, which Government departments/agencies may wish to use or adapt to meet their own Data Management needs.

Data acquisition

- All projects and other activities that give rise to substantial datasets will establish at the outset whether suitable data already exist in a potentially usable form, or whether new data need to be acquired.

- Before projects are approved, they must establish how the data acquired will be exploited to the full, who will be responsible for full exploitation of the data, and how the benefits will be maximized and shared.

- Subsequent data handling and storage needs will be considered, and plans put in place to ensure that databases are maintained in such a way that maximum use can subsequently be made of them.

Data care – Stewardship

- Databases will be managed closely, with clear responsibility for stewardship established and individuals made accountable for ensuring Data Management procedures are followed.

- Data will be held securely within their own database, and adequate provision made for their long-term care.

- All data will be validated and quality assured before being used or archived.

- Easy access will be given to data holdings, both for staff and bona fide ‘customers’.

- Data that are not legally required to be retained will not be destroyed or put at risk without first exploring all other possibilities and then demonstrating clearly that the costs of retaining them cannot be justified by potential benefits, or that the replacement cost is less than the storage costs.

Data use and exchange

- Memoranda of Agreement will be drawn up with Users and Customers who receive data, with respect to the subsequent use of such data. These will include confidentiality declarations and conditions of use.

- Intellectual Property Rights will be protected in relation to any development of information, by specifying formally any restrictions on the use of the data in formal licensing arrangements.

- Adequate provision will be made for the widest possible public access to data and associated metadata.

- Costs will be recovered for the handling of data and information, in line with departmental policies, which will be made readily available.

- The appropriate return will be charged when data are passed on to other parties seeking to make commercial gain.

Review

- The Data Policy will be monitored regularly and will be modified in the light of developments (e.g. technology and legislation) and experience. Information handling practices will be audited so that duplication can be minimized.

7 Implementation – key roles

To be successful, Data Management best practice must be implemented across the whole organization, under the guidance of a member of the Executive Board, i.e. the Data Management Champion. Other key roles are the Data Manager and the Data Stewards assigned to each key dataset.

The following list of responsibilities may help organizations to establish these key roles and implement good Data Management policies and procedures.

Data Management Champion

The Champion is responsible for:

- Ensuring that policies on Data Management are in line with legislation and Government Policies.
- Reporting progress to the Executive Board on the performance achieved against the targets set for the improvement of data quality and the value gained from effective Data Management.

In larger departments, particularly those spread over a number of sites, a Data Management Steering Group may also be required.

Data Manager

The Data Manager may require the help of Local Data Managers to discharge the following responsibilities:

- Developing and maintaining the Data Policy Statement and other corporate guidance.

- Directing the development, implementation and maintenance of the detailed data policies, standards, procedures and guidelines across the whole organization.

- Appointing and monitoring the performance of Data Stewards.

- Issuing guidance and training staff.

- Ensuring that local practice in individual business areas meets the standard set for the whole organization.

- Ensuring that the organization maintains a central metadata resource.

Data Stewards

Data Stewards are responsible for ensuring that the following minimum standards are applied for each dataset:

- The dataset must be documented in the organization's catalogue following the standards for discovery metadata, to enable the ownership, Intellectual Property Rights, stewardship and accessibility to be determined.

- The policy for exploiting the dataset and making it available to third parties must be agreed and documented.

- The dataset and its conditions of use must comply with all the statutory and non-statutory obligations of the organization.

- The data must follow standard classifications and definitions where appropriate, and must comply with all relevant standards, codes of practice and other protocols.

- The data must be fully validated and quality assured with sufficient detailed metadata to enable their use by third parties without reference to the originator of the data.

- The data must be stored, managed and accessed in line with agreed Data Management and Security/Confidentiality policies.

- The release/use of data by internal and external customers must be authorized and agreement to the conditions of use documented.

- The costs and benefits of continuing to maintain the dataset must be reviewed periodically.

8 Further guidance

This guide is intended to give an introduction to the principles of good Data Management and has been prepared with the help of a number of organizations who have already benefited from adopting such Data Management principles.

IGGI will continue to support the use of good Data Management principles and will, where possible, provide detailed guidance on the website. This detailed guidance will take the form of proven guidelines, made available by IGGI members. Comments on this guide and on the detailed guidance are welcomed, and these can be made to the IGGI Secretariat by e-mailing IGGI@odpm.gsi.gov.uk.

9 Glossary of Terms

AGI:	Association for Geographic Information.
Data Audit:	A process to demonstrate assurance that Data Management is being undertaken to the required level.
Data Management Champion:	The member of the Executive who is Corporately responsible for the Data Management.
Data Management Plan:	A plan for the management of an individual dataset, compliant with the local Data Policy.
Data Management Procedure:	A set of detailed operational parameters for the day-to-day management of a specific dataset.
Data Manager:	The senior manager, reporting to the Data Management Champion, responsible for Data Management in an organization.
Data Owners:	Are the individuals or groups of individuals who are held accountable, managerially and financially, for a dataset and who have legal ownership rights to a dataset even though that dataset may have been collected/collated/disseminated by another party.
Data Policy:	A set of broad, high-level principles that form the guiding framework within which Data Management can operate.

Data Quality:	A set of parameters, established according to specific local requirements and processes, to ensure that datasets may be seen to be fit for purpose.
Data Steward:	An individual accountable for the management of a specific dataset or group of datasets.
Data:	A collection of facts, concepts or instructions in a formalized manner suitable for communication or processing by human beings or by computer.
Dataset:	A dataset is a collection of data that has been compiled to serve a specific business purpose. It may have been collected for a specific purpose (e.g. Census data) or it may be data previously collected for another purpose that is being reused for a purpose not envisaged at the time of collection (e.g. family history research). Datasets are the fundamental unit of Data Management.
EIR	Environmental Information Regulations.
FoI	The Freedom of Information Act 2000.
Geographic information (GI):	Data/information that is referenced in some way to the earth's surface whether by co-ordinates or by geographic identifiers (addresses, administrative area, postcode).

IGGI:	The Intra-governmental Group on Geographic Information. Its mission is to "Increase the efficiency of central government while enabling it to meet its responsibilities for provision of geographic information to the general public".
IPR	Intellectual Property Rights.
Metadata:	Metadata is the term used to describe the summary information or characteristics of a set of data. In the area of geographic information or information with a geographic reference this normally means the What, Who, Where, When and How of the data. Thus, the only major difference that exists between this and the many other metadata sets being collected for libraries, academia, professions and elsewhere is the emphasis on the spatial component – or the Where element.
NGDF:	A now obsolete standard created by the former National Geospatial Data Framework. This was a name given to a public/private initiative. Its aim was to facilitate the unlocking of geographic information through enabling better awareness of data availability, improving access to the data and integrating data by encouraging the use of standards. Many of the functions of the NGDF are now fulfilled by the gigateway.
TNA	The National Archives.

10 World Wide Web URLs

[1] www.hmso.gov.uk/acts/acts2000/20000036.htm

[2] www.hmso.gov.uk/si/si2004/draft/20040331.htm

[3] www.hmso.gov.uk/acts/acts1998/19980042.htm

[4] www.hmso.gov.uk/acts/acts1998/19980029.htm

[5] www.nationalarchives.gov.uk/policy/act/act.htm

[6] www.nationalarchives.gov.uk/

[7] www.dca.gov.uk/

[8] www.govtalk.gov.uk/

[9] www.dca.gov.uk/foi/index.htm

[10] www.nationalarchives.gov.uk/policy/foi/

[11] www.nationalarchives.gov.uk/policy/environmental.htm

[12] www.fgdc.gov/publications/documents/geninfo/execord.html

[13] www.gigateway.co.uk

[14] www.iso.ch

[15] www.agi.org.uk

[16] e-government.cabinetoffice.gov.uk

[17] www.iggi.gov.uk/publications/index.htm

Geographic Information

IGGI:

The Intra-governmen

Geographic Informatio

"Increase the efficienc

ment while enabling

bilities for provision o