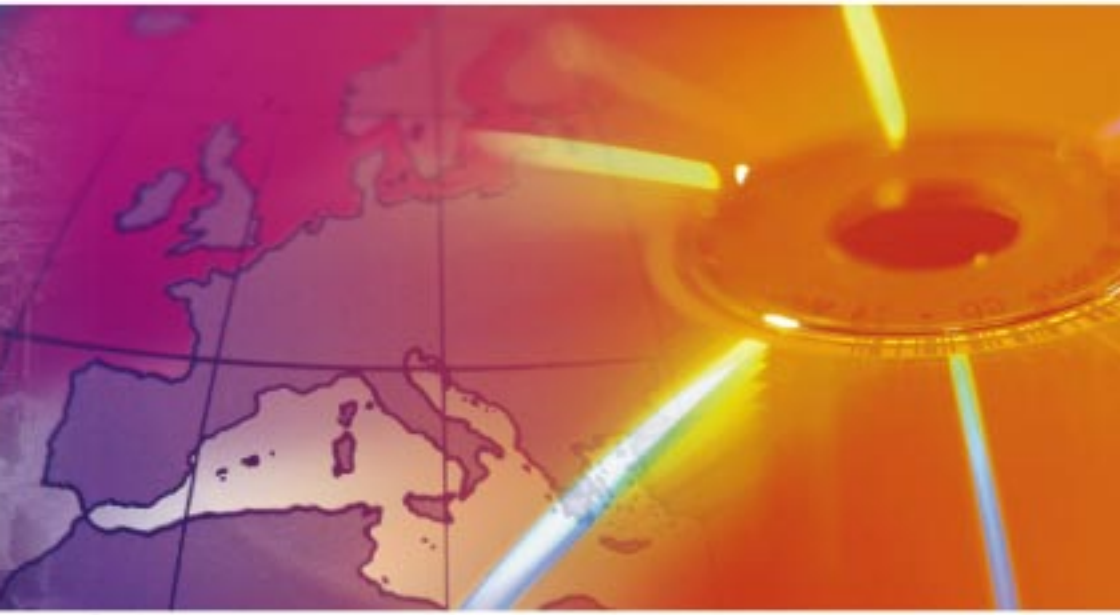
Intra-governmental Group
on Geographic Information



The Principles of Good Data Management

Intra-governmental Group on Geographic Information

The Principles of Good Data Management

This guide, version 1.0, September 2000 was prepared by the IGGI working group on Principles and Practice of GI Data Management. Thanks to Margaret Wagget (ONS) and her team, Stefan Carlyle (EA), James Denman (ONS), Jerry Giles (BGS), and Alan Oliver (DETR) for their considerable efforts in putting this guide together.

Further copies of this report are available from:
Department of the Environment, Transport and the Regions
IGGI Secretariat
3/H9 Eland House
Bressenden Place
London SW1E 5DU
Telephone 020 7944 5517
Fax: 020 7944 5519
e-mail: iggi@detr.gsi.gov.uk
Internet service http://www.detr.gov.uk/

© *Crown Copyright 2000*

Copyright in the typographical arrangement and design vests in the Crown.

Extracts of this publication may be made for non-commercial in-house use, subject to the source being acknowledged.

Applications for reproduction should be made in writing to The Copyright Unit, Her Majesty's Stationery Office, St Clements House, 1-16 Colegate, Norwich NR3 1BQ.

Published by the Department of the Environment, Transport and the Regions.
Printed in the UK, September 2000 on paper comprising 75% post-consumer waste and 25% ECF pulp.

Product code 00PD0835.

FOREWORD
By Alan Oliver, IGGI Chairman

One of the cornerstones of the Modernising Government agenda is better use of Information and Communications Technology for electronic delivery of services and information, for the benefit of business, the citizen and government itself.

Government is investing substantial resources in the development of ICT systems and networks. At the same time government policies are opening up government data for wider use. The technical barriers to electronic exchange of data have largely been overcome. However, the physical networks are only a means to an end, and the benefits of the new technology will only be realised if the information they carry is what users need.

Geographic Information (GI) provides a good example. Government departments and agencies hold huge amounts of geographic information, much of which has been collected for specific policy or operational purposes. Much of the value of GI lies in the ability to bring different data sets together for joint analysis. For this to be effective the data have to conform to recognised standards and delivery formats that the user can handle, and need to be accompanied by good documentation.

All of this adds up to a need for good data management.

Members of IGGI, whose aim is to improve the effective use of government Geographic Information, decided to pool their experiences and produce a self-help guide to good practice in data management. The result is this document, which will be distributed widely among holders of government geographic information. Since most of the principles apply equally to other types of data, we hope the guide may also be of wider value.

The guide sets out general principles but does not provide detailed descriptions for specific data management activities. Nor can it provide a full description of relevant government policies and initiatives. Data custodians should refer to other guidance for such detail. The Web version of this guide will be supported by references and links to other information. Comments and suggestions for improvement will be welcome – contact details are on the web site.

I would like to thank all those who contributed to the preparation of this guide, and hope that it can play a part in raising standards of data management in government.

CONTENTS

		Page
1.	**Purpose of this guide**	1
2.	**What is Data Management?**	1
3.	**Why do we need to manage our data?**	2
	Key drivers for improved Data Management	2
	Benefits of good Data Management	3
4.	**Principles of good Data Management**	5
	Data policy	5
	Data ownership	5
	Data documentation and metadata compilation	6
	Data quality, standardisation and harmonisation	7
	Data lifecycle control	8
	Data stewardship	8
	Data access and dissemination	9
	Data audit	9
5.	**Establishing a Data Policy**	10
	Data acquisition	10
	Data care – Stewardship	10
	Data use and exchange	11
6.	**Implementing Data Management – key roles**	11
	Data Management Champion	12
	Data Policy Manager	12
	Data Stewards	12
7.	**Further guidance**	13
	Annex 1. Glossary of Terms	14

1. Purpose of this guide

Government departments and agencies collect, generate, store and use large amounts of data which have been obtained at considerable cost. Many of these data are geographical in that they are referenced to geographical locations, such as points, lines or areas.

The booklet, entitled "*Geographic Information: A Charter Standard Statement*" published by the Intra-governmental Group on Geographic Information (IGGI) in September 2000, sets out the standards of good practice on information handling that users can expect. Many individual departments and agencies have used this broad framework as the basis for their own geographic information policy statement.

This guide provides the next step in advising officials responsible for information handling, by providing general guidance on managing data as a valued resource. The guide had been produced for those responsible for Geographic Information, although the management principles are equally relevant to other types of government data.

2. What is Data Management?

Data Management embraces the whole range of activities involved in the handling of data. These activities, include: Data Policy; Data Ownership and responsibilities for ensuring Legislative Compliance; Data Documentation and Metadata Compilation; Data Quality, Standardisation and Harmonisation; Data Lifecycle Control; Data Stewardship; Data Access and Dissemination; and Data Audit.

This guide covers only the key aspects of Data Management. Other more detailed guidance can be found on the IGGI Web site (www.iggi.gov.uk).

3. Why do we need to manage our data?

Government owns huge amounts of irreplaceable Geographic Information potentially usable by a wide range of users 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 strong direction from the Modernising Government programme of work, and particularly the strategy for Information Age Government, to improve services and make more effective use of Government information resources.
- Increasing recognition that Government data, collected at public expense, must be properly managed in order to realise 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.
- Stronger emphasis within Government on the need to rationalise and combine data in order to improve efficiency and add value.
- More reluctance from suppliers to provide data at affordable prices. Stricter control required by Data Owners over the use of their data to safeguard their Intellectual Property Rights (IPR) and the confidentiality of sensitive data.
- The increasing number of legal and political directives and other obligations which impact on the management of data and the service to customers, including European directives, Government policies and departmental guidance. Examples include the Data Protection and Freedom of Information Acts, the Public Records Act, Crown Copyright, the “Building Trust In Statistics” white paper.

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

- 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.
- Clear understanding of the subsequent use of their data, documented in a formal Memorandum of Agreement signed by supplier and user.
- Fair return for the use of their data.

Benefits to Data Brokers/Intermediaries

- Better quality, harmonised 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 effective data policies and best practice guidance on business justification, documentation, maintenance, data audit, archiving and final destruction, where appropriate.
- 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 standardisation of datasets that are frequently used by different parts of an organisation.

- Increased confidence that the organisation complies with statutory and non-statutory obligations, by the regular use of centrally co-ordinated, 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 organisation 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 which recognise the need for free access by the appropriate customers whilst recovering the appropriate income from customers who seek to make commercial gain.
- Increased trust and confidence of customers in the quality and reliability of outputs.

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 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 effective metadata.
- Better value for money, resulting from clear, fair and consistent data charges and conditions of use, which recognise the need for free access by the appropriate customers.
- Better exploitation of data generally, enabled by easier data exchange and integration with other harmonised data.
- Efficiency gains resulting from easier use of better quality data.

4. Principles of good Data Management

Section 2 has already identified the main activities involved in good Data Management. This section provides an outline of each of these key activities. Further guidance can be found on the IGGI website.

Data policy

The first step for any organisation wishing to implement good Data Management procedures is to define a Departmental Data Policy. This is a set of broad, high-level principles which form the guiding framework in which Data Management can operate. Section 5 contains further information, including 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 organisation or group of organisations who originally commissioned the data and has managerial and financial control of that data. The Data Owner has legal rights over the data, the Intellectual Property Rights and the Copyright. This is still the case, even where the data were collected, collated or disseminated by another party.

Data ownership implies the right to exploit the data, and if continued maintenance becomes unnecessary or uneconomical, the right to destroy them. Ownership can relate to a data item, a merged 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 organisation, even though other organisations 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.

Data documentation and metadata compilation

All datasets should be identified and documented to facilitate their subsequent identification, proper management and effective use, and to avoid collecting or purchasing the same data more than once.

To provide an accurate list of datasets held by the organisation, a departmental catalogue of data should be compiled. This is a collection of discovery level metadata for each dataset, in a form suitable for users to reference. These metadata should provide information about the content, geographic extent, currency and accessibility of the data, together with contact details for further information about the data.

All business-related datasets, once catalogued, should also be documented in a detailed form suitable for users to reference when using the data. These detailed metadata should describe the content, characteristics and use of the dataset, using a standard detailed metadata template.

Further guidance on the compilation of metadata, including the two templates mentioned above, can be found on the IGGI website. Discovery metadata collection and maintenance tools and information on the Data Locator service can be found on the National Geospatial Data Framework (NGDF) website (www.ngdf.org.uk)

Data quality, standardisation and harmonisation

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.

To maximise the potential and use of datasets, organisations should:

- Use standard data definitions and formats.
- Define quality standards and apply the appropriate validation processes to each dataset.
- Adopt formal Query and Change Management procedures.
- Ensure that data are quality assured and approved as fit for purpose before use or release.
- Encourage the use of the appropriate National and International standards, and particularly those which are relevant to Geographic Information.

The UK Standard Geographic Base (UKSGB) is one initiative which aims to improve the quality of geographic information by providing easy access to standard set of geographic base data for commonly used areas in the UK, for example administrative, electoral, census and postal areas. Further information can be found on the IGGI website.

Data lifecycle control

Good Data Management requires that the whole lifecycle of data be managed carefully. This includes:

- Business justification, to ensure thought has been given to why new data are required rather than existing data amended, 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.

Data stewardship

There are a number of focal points within Government organisations where data are gathered, compiled and analysed as part of normal course of business. These focal points have a particular role to play in managing the Government's data, over and above the responsibilities of Data Owners.

An appointed individual, called a Data Steward (or in some organisations a Data Custodian) should be given formal responsibility for the stewardship of each major dataset. The Data Steward should be made responsible and 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 access and dissemination

Although this aspect will depend on the business and the financial policy of the organisation, the following guidance may be helpful.

- Public access to data should be provided in line with the Open Government code of practice, the Freedom of Information Act and the Data Protection Act. See IGGI website.
- Access to data should be granted to customers and commercial organisations when the request is in line with the organisation's business strategy and is not infringing the Copyright/IPR of the data or any statutory/departmental obligations.
- The right to use or provide access to data can be passed to a third party subject to agreed pricing and dissemination policies.

Data audit

Data management audits are recommended to assess the extent to which Data Management procedures are followed. The following mechanisms may be used to monitor implementation.

- Data Management audits of major data collection, storage and dissemination activities should be commissioned to ascertain the level of compliance with data policies and guidance notes.
- Data management audits should review the extent to which stewardship procedures are followed and the subsequent improvement in the quality and accessibility of the data is subsequently improved.
- Memoranda of Agreements between owner/suppliers and users should be checked to ensure that the conditions of use are in line with agreed departmental policies.

5. 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, which give rise to substantial data sets, will establish at the outset whether suitable data already exists in a potentially usable form, or whether new data needs 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 maximised 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 stewardship procedures are followed.
- Data will be held securely on 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, which 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 Agreements 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.

6. Implementing Data Management – key roles

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

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

Data Management Champion

The Champion is responsible for:

- Ensuring that departmental policies on Data Management are in line with Government Policies.
- Directing the development, implementation and maintenance of the detailed data policies, standards, procedures and guidelines across the whole organisation.
- 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 Policy Manager

The Data Policy 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.
- Appointing and monitoring the performance of Data Stewards.
- Issuing guidance and training staff.
- Ensuring local practice in individual business areas meets the standard set for the whole organisation.
- Ensuring that the organisation 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 organisation'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 statutory and non-statutory obligations of the organisation.
- 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 authorised and agreement to the conditions of use documented.
- The costs and benefits of continuing to maintain the dataset must be reviewed periodically.

7. 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 organisations 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 on the IGGI website (www.iggi.gov.uk).

Glossary of Terms

Data:
A collection of facts, concepts or instructions in a formalised manner suitable for communication or processing by human beings or by computer.

Data Owners:
Are the individuals or groups of individuals who are held managerially and financially accountable for a Data set and who have legal ownership rights to a Data set even though that Data set may have been collected/collated/disseminated by another party.

Data Policy:
A set of broad, high-level principles which form the guiding framework in which Data Management can operate.

Data Sponsors:
Are individuals or groups of individuals who commission the collection, compilation or production of Datasets. Data Sponsors may, in addition, be the Owners of such Datasets especially if they provide the funding for such collection or production efforts.

Gateway:
An interface between some external source of information and a World Wide Web server. In this instance a gateway is a web enabled search mechanism which allows users to search a distributed network of directory nodes.

Geographic information:
Data/information that is referenced in some way to the earth's surface whether by co-ordinates geographic identifiers (addresses, administrative area, postcode).

GICSS:
The Charter Standard Statement is designed to promote the effective use of geographic information held by government. The statement provides a broad framework, within which individual departments and agencies may issue their own geographic information policy statements.

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".

Information Age Government:
Exploiting Information and Communication Technologies to Transform Public Services.

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. The only major difference that therefore exists from 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:
The National Geospatial Data Framework is the name given a public/private initiative. Its aim is 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.

The Principles of Good Data Management guidelines:
This guide provides officials responsible for information handling general guidance on managing data as a valued resource. The guide refers to data with a geographic content, although the management principles are equally relevant to other types of government data.

Geographic Information

IGG1:

The Intra-governmen

Geographic Informatio

"Increase the efficiency
ment while enabling it
bilities for provision of