WCMC Handbooks on Biodiversity Information Management

A DARWIN
INITIATIVE PROJECT

WORLD CONSERVATION
MONITORING CENTRE

COMMONWEALTH
SECRETARIAT

Companion Volume

WCMC Handbooks on Biodiversity Information Management

Companion Volume

World Conservation Monitoring Centre

**WORLD CONSERVATION
MONITORING CENTRE**

Series Editor J.H. Reynolds

Commonwealth Secretariat   1998

WORLD CONSERVATION
MONITORING CENTRE

Contents

ACKNOWLEDGEMENTS

editor for the handbooks was Jake Reynolds, while Donald Gordon managed the overall project.

To the many individuals, both within and outside WCMC who have contributed to the development of materials and the delivery of training in biodiversity information management, a profound debt of gratitude is owed. It is through this collaborative effort that a service is being developed to contribute to the conservation and sustainable use of living resources.

BACKGROUND

The purpose of the *WCMC Handbooks on Biodiversity Information Management* is to support those making decisions on the conservation and sustainable use of living resources. The handbooks form part of a comprehensive programme of training materials designed to build information-management capacity, improve decision-making and assist countries in meeting their obligations under Agenda 21 and the Convention on Biological Diversity.

The intended audience includes information professionals, policy-makers, and senior managers in government, the private sector and wider society, all of whom have a stake in the use or management of living resources. Although written to address the specific need for improved management of biodiversity-related information at the national level, the underlying principles apply to environmental information in general, and to decision-making at all levels. The issues and concepts presented may also be applied in the context of specific sectors, such as forestry, agriculture and wildlife management.

The handbooks deal with a range of issues and processes relevant to the use of information in decision-making, including the strengthening of organisations and organisational linkages, data custodianship and management, and the development of infrastructure to support data and information exchange. Experience suggests that some of the greatest challenges in information management today are concerned with organisational issues, rather than technical concerns in the delivery of information which supports informed decision-making. Consequently, topics are addressed at management and strategic levels, rather than from a technical or methodological standpoint, and alternative approaches are suggested from which a selection or adaptation can be made which best suits local conditions. Nevertheless, in adopting this framework approach, we have tried to adhere to recognised conventions and formalisms used in information management and trust that in producing a 'readable' set of handbooks the integrity of the materials has not been compromised.

Overall, the handbook series comprises:

Companion Volume
Volume 1 Information and Policy
Volume 2 Information Needs Analysis
Volume 3 Information Product Design
Volume 4 Information Networks
Volume 5 Data Custodianship and Access
Volume 6 Information Management Capacity
Volume 7 Data Management Fundamentals

Collectively, the handbook series promotes a shift from tactically based information systems, aimed at delivering products for individual project initiatives, to strategic systems which promote the building of capacity within organisations and networks. This approach not only encourages data to be managed more effectively within organisations, but also encourages data to be shared amongst organisations for the development of the integrated products and services needed to address complex and far-reaching environmental issues.

The handbook series can be used in a number of ways. Individual handbooks can be used to guide managers on specific aspects of information management; they can be used collectively as a reference source for strategic planning and project development; they can also provide the basis for a series of short courses and training seminars on key challenges in information management.

The companion volume provides the background to the handbook series. It also assists readers in deciding which handbooks are most relevant to their own priorities for strengthening capacity.

A second series of handbooks is planned to provide more detailed guidance on information management methodologies, including the areas of data and technology standards, database design and development, application of geographic information systems (GIS), catalogues and metadatabases, and the development of decision-support systems. The current series deals only briefly with formal system development methodologies, and for more detailed treatments the reader is encouraged to access the wide range of published and electronic resources available in libraries and on the Internet, some of which are alluded to in individual handbooks and reference sections.

A number of computer-based training tools have been developed to accompany the handbook series and are used in the training programme. These are based on a protected areas database, a tree conservation database, a GIS demonstration tool and a metadata directory. They aim to demonstrate key aspects in the collection, management and analysis of biodiversity data, and the subsequent production and delivery of information. They also illustrate practical issues such as data standards, data quality-assurance, data access, and documentation. Each training tool is supported by a user guide, together with a descriptive manual which traces the evolution of the tool from design, through development to use.

1   INTRODUCTION

Over the past three decades, perceptions of the environment have gone through a number of shifts, often marked by a series of landmark events. For example, the myth of 'infinite resources', which pervaded decision-making until the 1960s, gradually gave way to an era of 'environmental protection', which focused on the environmental impacts of waste production and over-consumption. This was followed in the 1970s and '80s by a 'resource management' era, which took a sectoral approach to environmental management. The paradigm of the 1990s is 'sustainable development', or the integration of environmental, social and economic goals in an holistic, equitable mix.

Landmarks in this progression include the United Nations Conference on the Human Environment, also known as the Stockholm Conference of 1972, which spawned institutional mechanisms to address global environmental issues, including the United Nations Environment Programme; the World Conservation Strategy in 1980, linking conservation and development issues; the World Commission on Environment and Development (WCED) in 1987, which gave international recognition to the notion of sustainable development; and the United Nations Conference on Environment and Development (UNCED), also known as the Earth Summit, in 1992.

Over this time, increased recognition has been given to the management of information for environmental decision-making, as well as to the strengthening and networking of organisations. These priorities are vitally important to the development of national strategies, plans and programmes in support of international treaties and conventions on living resources, and are reflected both in Agenda 21 and in the articles of the Convention on Biological Diversity.

Against this background, the Darwin Initiative project entitled *Training in Biodiversity Information Management* was approved in 1996. The Darwin Initiative programme has to date supported more than one hundred and forty projects at a total cost of £15 million. This project supports three of the five principal areas on which the Darwin Initiative focuses, namely institutional capacity building, training, and assisting in the implementation of the Convention on Biological Diversity.

2   THE HANDBOOK SERIES

The objective of the handbook series is to support policy and management goals aimed at conserving and sustainably using living resources. The handbooks are based upon the following three principles:

- Information needs to be timely and available in forms which genuinely support decision-making.

- Data should be collected and quality-assured once, and then made available for use in multiple information products.

- Policy-makers, resource managers, information professionals and researchers should attempt to understand each other's perspectives and cooperate in the production of information.

To fulfil these ideals it is necessary to develop an infrastructure capable of storing relevant data, harnessing professional skills to generate information, and then using this information to influence major stakeholders in favour of conservation goals. Ultimately, the last of these is the only true measure of success.

A key requirement is to respond to environmental concerns before they become serious problems. Most countries have some capacity to do this already, although the capacity may not be well coordinated or funded. Remembering that most systems are only as strong as their weakest link, a secondary objective of the handbook series is to help readers judge how far their infrastructure has progressed and identify areas which would benefit from further attention and investment. Although the handbooks are closely interrelated, each has its own individual focus as illustrated in Figure 1 and described below.

Figure 1   The focus of each handbook
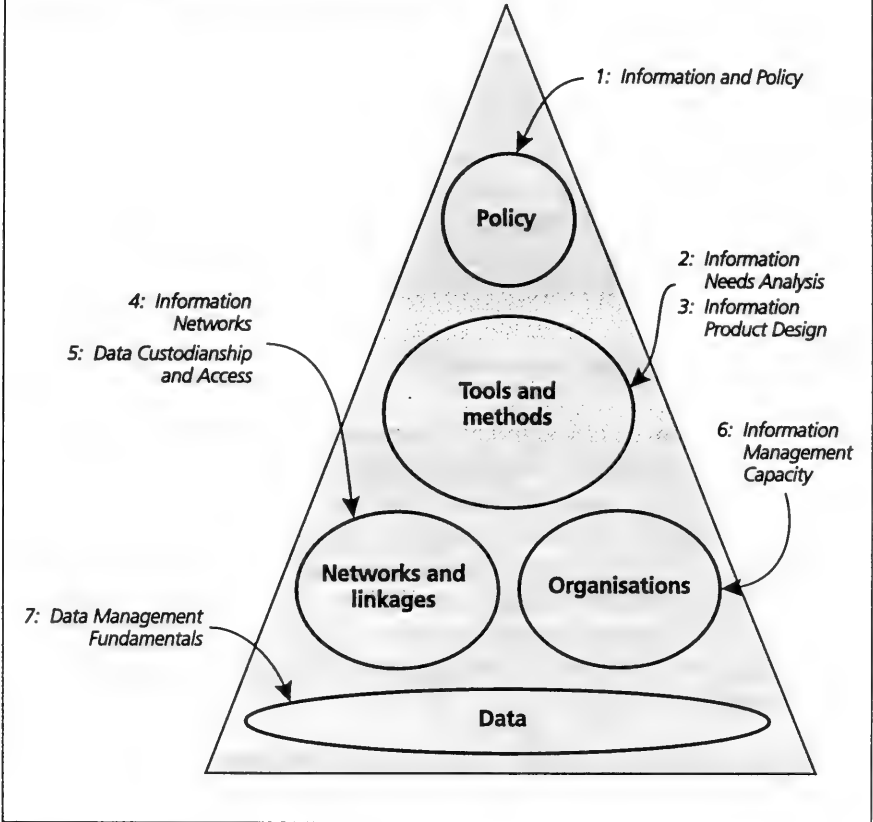
- **Volume 1   Information and Policy**

This handbook examines the links between information management and policy, in particular the use which is made of scientific information in the policy-development process. The handbook introduces models for how information can be applied effectively to decision-making (the 'management loop'), and how information professionals, policy-makers and resource users can cooperate in generating information to address environmental concerns (the 'information cycle'). Case studies are drawn from India and the UK to illustrate

the connection between information and policy development. Key topics covered in the handbook include:

✓ setting priorities for information management

✓ participation and consensus

✓ supporting good decisions

✓ supporting good policies

✓ the concept of an 'information cycle'.

● **Volume 2   Information Needs Analysis**

This handbook provides an outline of the rationale and benefits of conducting an information needs analysis to ensure that investments in information are valued by key decision-making groups. The analysis involves identifying key stakeholders (users); determining their goals with respect to specific issues; identifying constraints on information usage; and determining what information is needed to implement, monitor or review given policies. The underlying goal of the analysis is to achieve a consistent, mutually agreed set of information priorities amongst stakeholders, and to integrate the perspectives of users into a common vision of the way forward.

The handbook outlines some practical approaches which may be used to identify the information needs of decision-makers. Finally, a case study is provided on how an information needs analysis was carried out in the forest sector of a tropical country. Key topics in the handbook include:

✓ why an information needs analysis is necessary

✓ how information needs are determined

✓ tools and methods which can assist in the process.

• Volume 3   Information Product Design

This handbook reviews the design and development of information products with consideration given to content, complexity, structure and layout, and timing in the delivery. The handbook also reviews data needs, underscoring the importance of developing essential datasets for multiple purposes. A review of processing needs considers steps such as data collection, data storage/quality assurance, interpretation/analysis, through to the compilation and presentation of information products. The handbook then considers the development and marketing of strategic products, potentially involving a range of cooperating partners, and introduces a case study on the *World Forest Map*. Key topics include:

✓ features of a good information product

✓ policy-relevant information

✓ abstraction, summary and indicators

✓ analysing data and processing needs

✓ product development

✓ packaging and marketing information.

• Volume 4   Information Networks

This handbook outlines the principles of information networks and explores centralised, distributed and managed networks which largely reflect evolutionary developments in information sharing and access. Central to the effective operation of information networks are clearly defined roles and responsibilities of the partners, i.e. custodians, users and the 'hub'. Partners may, however, play multiple roles and be part of network arrangements operating across different themes and at different levels, from local through to international.

Successful networks have a number of features in common: they have a recognised and easily understood purpose; there is an effective dialogue with users; they are flexible to changing circumstances and demands; and they mobilise and build on existing resources and capabilities in terms of the data, skills and

facilities of a range of partners in support of common information objectives. Four case studies of information networks developed at the global, regional (East Africa) and national levels (Australia and the UK), respectively, accompany this handbook. Key topics include:

✓ networking principles

✓ network design

✓ partner roles and responsibilities

✓ achieving common objectives

✓ network sustainability.

● **Volume 5   Data Custodianship and Access**

This handbook deals with the principles of custodianship of datasets, drawing distinction between custodians, owners and users. The rights and responsibilities of custodians are outlined, as are considerations involved in the management of custodianship, including assigning, managing and reviewing custodial responsibilities. A technique is outlined to assist steering committees in the determination of custodianship, and the handbook provides guidance on the development and use of data access agreements. Quality-assurance and access to essential datasets are fundamental principles, while pricing strategies are also considered. The *Commonwealth Custodianship Guidelines*, from Australia, provide a useful case study of the real life application of custodianship. Key topics include:

✓ principles of custodianship

✓ functions of custodians

✓ assigning and managing custodianship

✓ data access agreements.

- **Volume 6   Information Management Capacity**

  Every organisation possesses its own individual strengths, weaknesses and objectives for information management. This handbook examines how to assess the current level of capacity within an organisation, with particular reference to direct assets (e.g. datasets, expertise, facilities) and indirect assets (e.g. management systems, partnerships), providing the basis for strategic planning and investment in capacity building.

  The handbook outlines components of an institutional survey, supported by a comprehensive questionnaire as an annex, and reviews the analysis and presentation of survey results. An overview of strategic planning is provided, and the handbook considers the key steps in enhancing an organisation's information management capacity, particularly as it supports wider goals within networks. A case study on national institutional survey in Ghana is then introduced, with reference to strategic planning and policy support. Key topics include:

  ✓ assessment of capacity

  ✓ suggestions for analysis

  ✓ strategic planning

  ✓ capacity building.

- **Volume 7   Data Management Fundamentals**

  This handbook deals with the development of an efficient infrastructure to underpin information production. The aim is to increase the usefulness of datasets by ensuring that they are flexible, standardised, of a high quality and accessible. A basic premise is that primary data should be managed as the basis for the production of a wide range of information products.

  The handbook provides a discussion of data standards, validation and maintenance of data, documentation, data security, and the use of quality-management standards. Consideration is given to the selection of information technology, particularly as it relates to the concepts of scalability,

connectivity, compatibility and sustainability. The handbook concludes with a section on database development.

A *Tree Conservation Database* serves as a case study, illustrating many aspects of the information cycle introduced in Volume 1 (i.e. from identification of information needs to the establishment of an information service underpinned by data which are managed following accepted standards and information management procedures). The Tree Conservation Database also serves as one of the complementary training aids to the handbook series. Key topics include:

✓ data flexibility

✓ data standards

✓ quality-assurance

✓ information technology

✓ database development.

3 WORKING DEFINITIONS

Working definitions of the following terms are provided to help users understand the text of the handbooks, and as a general reference source. No claim is made that these definitions are universally accepted or authoritative.

Application   Software fulfilling a specific function on a computer, or an equivalent manual procedure. Can be general-purpose (e.g. a word-processor) or custom-built to undertake specific tasks.

Attribute   Properties of an entity (e.g. 'name' and 'area' might be attributes of the 'Protected Areas' entity).

Audience   The target user (user group) for an information product.

Benchmark   A measure of performance against a specific target value or values.

Best-practice   Voluntary methods of working which are widely recognised as ethical, sustainable and otherwise sound.

Biodiversity   The variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are a part; this includes diversity within species, between species and of ecosystems. The term is used interchangeably with biological diversity.

Biological resources   Includes genetic resources, organisms or parts thereof, populations, or any other biotic component of ecosystems with actual or potential use or value for humanity.

Bulletin board   An area set aside in an electronic communications network where messages can be posted and read by a restricted group of members. Also known as a 'newsgroup'.

CD-ROM (Compact Disc-Read Only Memory)   Optical technology which enables large amounts of data to be stored digitally on an optical disk in read-only format.

Client-server A computing architecture which offers the advantages of both stand-alone and fully networked architectures. Computing tasks are shared efficiently between a high-powered server computer and individual user's workstations.

Computer network A group of computers which can communicate with each other.

Continuous improvement A management paradigm in which long-term goals are achieved through a succession of small, individually contributing steps, with frequent review.

Data Facts that result from measurements or observations of a phenomenon.

Data development The process of building from scratch or upgrading a dataset.

Data dictionary A repository of information about the definition and structure of a database.

Data-flow diagram A diagram illustrating how information and data flow in a process (e.g. in an organisation). Special symbols depict different types of flow.

Data management The organisation, storage, retrieval and maintenance of data in a controlled environment (see information management).

Data quality A statement of the quality of the data from which, a datasets fitness for use for a specific purpose may be assessed.

Data standards A standard set of terms. Documented norms, rules, conventions and guidelines concerning the data — includes data collection methodologies, definitions, coding and classification rules, representation conventions (units of measure, scale, precision . . .), allowable values and ranges.

Database A logically structured and consistent set of data, usually managed as computerised files in one or more locations.

Database application A collection of tools which facilitates use of a database.

Database Management System (DBMS)   A software package for data management that organises data in a defined structure that is independent of the application programs, and allows multiple users to share common data. DBMS software will normally include a data definition language, and data manipulation language.

Dataset   A collection of data and accompanying documentation which relate to a specific theme.

DBF format   The file format originally used by the dBASE product and now commonly used by other applications for data exchange.

Decision-maker   An individual or group responsible for making decisions (which impact on the well being or management of biological resources).

Decision-making   The process of arriving at decisions, which may be very complex and involve many organisations and individuals.

Digitising table   A device for inputting map features into a computer.

Electronic mail (email)   A network (including Internet) service allowing messages and files to be exchanged between users.

Entity   Physical objects or phenomena that possess attributes or properties (quantitative or qualitative) and may be considered to have relationships to other entities. For example: protected area, soil unit, species, manager, ecosystem and river.

Entity-Relationship (E-R) diagram   A graphic representation following a mathematical formalism that defines the inter-relationship between entities in a database.

Essential dataset   A dataset which underpins the development of multiple information products for multiple users. Examples include the national boundary, vegetation coverage and human population statistics.

Field   A column in a database table containing data values from a consistent data collection or interpretation process.

File   An organised collection of related records.

File Transfer Protocol (FTP)   A standard by which a user can access and transfer a file from one host (e.g. computer) to another using an FTP application.

File server   A specialist computer designed to manage access to shared data storage and management facilities.

Flat-file database   A simple type of database containing only one table of data ('flat-file').

Geographic Information System (GIS)   A software package that enables users to store and analyse spatial data, such as digital maps and remotely sensed imagery.

Global Positioning System (GPS)   A device which uses satellite signals to calculate the absolute or relative location of a point on the Earth's surface. Used in aircraft, ships and, increasingly, by biological survey teams, for instance to locate sampling or specimen collection points.

Hard copy   Data or information that is printed on paper.

Hardware   The physical components of an information system, such as computers, printers, scanners, plotters, and other peripheral devices.

Hyperlink   The connections between pages in a hypertext document.

Hypertext   Documents which are structured into electronic 'pages' connected to one another by means of hyperlinks.

Hyper Text Markup Language (HTML)   A standard language used to create hypertext documents, for instance to format pages displayed using web browsers.

Indicator   Highly refined information product which quantifies or simplifies complex phenomena, for example a numeric value representing the relative greenhouse gas emissions from different countries.

Information   Information is derived from the assembly, analysis or summarising of data into a meaningful form. Note: In some common usage there is often no clear distinction made between 'data' and 'information'. In information systems terminology it is common to use 'data' for the input to any process and call the output 'information' — which may then subsequently be the 'data' that is input into the next process.

Information management   The transformation of data into information, often through a set of processes involving people, equipment and procedures (see data management).

Information network   A group of partners/stakeholders collaborating with each other for the purposes of generating and exchanging information.

Information product   One or more items of information designed for a specific audience for a specific purpose (includes reports, graphs, documents, maps, multimedia presentations, for example).

Information production   The process of developing, packaging and communicating information products.

Information professional   Person involved with information production, for example a researcher, data manager, librarian, statistician, analyst, designer, publisher or communicator.

Information service   A facility offering a range of information products in support of a particular topic (e.g. forestry).

Information system   An organised set of people, processes, data and tools for transforming data into information and providing and disseminating information products.

Internet   A network of computers around the world which communicate using a set of agreed protocols (communication standards). The Internet provides useful services such as email, World Wide Web (WWW), Gopher and File Transfer Protocol (FTP).

Living resources   See biological resources.

Listserver   An Internet service based on simple email technology, which enables defined groups of users to be reached with a single email message. Primarily used for discussion of a specific topic (similar to bulletin boards).

Local Area Network (LAN)   A computer network usually operating within a single site or institution.

Logical database design   The conceptual design of a database which is independent of implementation issues.

Mainframe   A powerful multi-user computer designed to meet the demands of large organisations. Mainframes have been increasingly displaced in recent years by personal computers, workstations and communications networks.

Metadata   Literally, data about data (e.g. its location, source, content and quality) Also known as co-data.

Metadatabase   A database designed to manage metadata.

Modem   A device used to link computers over telephone lines. The term is a contraction of modulator-demodulator.

Multimedia   The integration of many types of data in a single application, including text, sound, graphics, and video.

Multitasking   A computing environment that enables several programs to be run concurrently.

Natural resources   Naturally occurring commodities used by human beings, for example minerals, soil, water and living resources such as forests, domesticated plants and animals, and wildlife. Renewable natural resources are a subset which regenerate under the right conditions.

Normalisation   The process of achieving an optimum structure for a relational database.

On-line database   An information service that can be accessed from computers dialling up over public communication networks.

Operating system   Software controlling access to the resources of a computer, including supervision of other program. Examples of operating systems are Microsoft Windows and UNIX.

Personal Computer (PC)   A computing platform, intended for use by one person at any time. Can be used either as a stand-alone computer or linked into networks. Major types include IBM-PC Compatible and Macintosh.

Physical database   A database which has been implemented in a particular hardware or software configuration.

Policy   A principle adopted or pursued by an individual, government, party or business.

Process   An activity, function or procedure applied to a resource to create some outcome (e.g. an arithmetic procedure applied to data, or a critical step in a business operation).

Prototyping   An information system development methodology which develops a partial or 'mock-up' version of an information system for user evaluation. Prototypes may be discarded or subsequently developed into an operational version.

Public domain   Information (e.g. software, databases, methods, standards) which can be used without restriction of copyright, licence and the like is said to be 'in the public domain'.

Query   A request to a database to select and extract data.

Record   A collection of related data about a specific topic, treated as a single unit for purposes of data management or analysis.

Relational database   A database in which data is organised following the 'relational model', that is, data items are organised in a set of formally-described tables linked by common key data items.

Relational Database Management System (RDBMS)   A DBMS which uses a relational data model as its basis.

Relationship   Describes how two or more entities are related to one another (e.g. 'species' may be related to 'genera' by a 'belongs to' relationship).

Server   A computer or program that provides a service to other programs or users. A network server, for example, enables users to access network resources from linked computers or terminals.

Software   The programs that are run on a computer.

Spatial data   Data which are associated with specific locations on the Earth's surface. Also known as geo-referenced data.

Spreadsheet   Software which allows users to manipulate and analyse data in tabular format, including the use of formulae to derive values in "cells".

Stakeholder   Individuals or groups having an interest in the well-being or management of a resource, for example, national and local government agencies, non-governmental organisations, community-based organisations, the private sector, industry, the general public, politicians, individuals and the international community.

Structured Query Language (SQL)   Widespread database querying language used in many relational database packages.

Sustainable use   Means the use of components of biological diversity in a way and at a rate that does not lead to the long-term decline of biological diversity, thereby maintaining its potential to meet the needs and aspirations of present and future generations.

Table   A physical entity in a relational database in which data are laid out in rows (records) and columns (fields).

Theme   A broad data area which may be subdivided into datasets.

Wide Area Information Server (WAIS)   A text-based search facility designed for retrieving information from computer networks.

Wide Area Network (WAN)   A computer network consisting of geographically dispersed computers communicating via media such as telephone, radio and satellite.

Workstation High-performance desktop computer designed for intensive technical applications.

World Wide Web (WWW) An Internet service enabling users to access information via a graphical, hypertext, interface.

Universal Resource Locator (URL) A unique address describing the location of an information source on the Internet.

xBASE Data management applications which trace their origins to the dBASE package.

WORLD CONSERVATION
MONITORING CENTRE

WCMC Handbooks on Biodiversity Information Management

Installation:

- In Windows Explorer, double-click on the file *a:\handbook*
- Choose a directory (C:\HANDBOOK). Select Unzip® to install 8 Adobe Acrobat PDF files. Click on OK, then Close.
- Open these files in *Adobe Acrobat Reader*. If you do not have this *Reader*, download it from the following web page: *http://www.adobe.com/prodindex/acrobat/readstep.html*

© 1998 World Conservation Monitoring Centre
219 Huntingdon Road, Cambridge CB3 0DL, UK. Tel: +44 1223 277314, Fax: 277136

WCMC Handbooks on Biodiversity Information Management

These handbooks have been developed for use by senior decision-makers and mid-career professionals. They review the issues and processes involved in the management of biodiversity information to support the conservation and sustainable use of living resources. They also provide a framework for the development of national plans and strategies and for meeting reporting obligations of international programmes and conventions. Collectively, the handbook series may be used as a training resource or, more generally, to support institutions and networks involved in building capacity in information management.

Companion Volume
Volume 1 Information and Policy
Volume 2 Information Needs Analysis
Volume 3 Information Product Design
Volume 4 Information Networks
Volume 5 Data Custodianship and Access
Volume 6 Information Management Capacity
Volume 7 Data Management Fundamentals

DARWIN INITIATIVE SECRETARIAT
4-A2 Ashdown House, 123 Victoria Street
London SW1E 6DE. United Kingdom

COMMONWEALTH SECRETARIAT
Marlborough House, Pall Mall
London SW1Y 5HX. United Kingdom

WORLD CONSERVATION MONITORING CENTRE
219 Huntingdon Road
Cambridge CB3 0DL
United Kingdom

Tel: +44 1223 277314
Fax: +44 1223 277136

E-mail: info@wcmc.org.uk

WCMC Internet Home Page
http://www.wcmc.org.uk