



A DARWIN
INITIATIVE PROJECT



WORLD CONSERVATION
MONITORING CENTRE



COMMONWEALTH
SECRETARIAT

Data Custodianship and Access



Volume 5

Digitized by the Internet Archive
in 2010 with funding from
UNEP-WCMC, Cambridge

<http://www.archive.org/details/wcmchandbooksonb05reyn>

WCMC Handbooks on Biodiversity Information Management

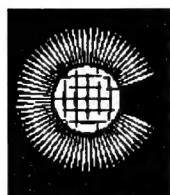
Volume 5
Data Custodianship and Access

World Conservation Monitoring Centre



**WORLD CONSERVATION
MONITORING CENTRE**

Series Editor J.H. Reynolds



Commonwealth Secretariat 1998

The **World Conservation Monitoring Centre**, based in Cambridge, UK, is a joint venture between three partners in the *World Conservation Strategy* and its successor *Caring for the Earth*: IUCN – The World Conservation Union, UNEP – United Nations Environment Programme, and WWF – World Wide Fund for Nature. The Centre provides information services on the conservation and sustainable use of species and ecosystems and supports others in the development of their own information systems.

The United Kingdom's **Darwin Initiative for the Survival of Species**, launched at the 1992 Earth Summit in Rio de Janeiro, aims to support the Convention on Biological Diversity by drawing on Britain's scientific, educational and commercial strengths to assist in the conservation and sustainable use of the world's biodiversity and natural habitats. Key tenets of the Darwin Initiative include collaboration and cooperation with local people, capacity building, distinctiveness and complementarity of project initiatives, poverty alleviation, and long-term sustainability. Through training, awareness raising, and research on undervalued areas of biodiversity, Darwin support is particularly aimed at strengthening links between Britain and those countries rich in biodiversity but poor in financial resources.

Under the auspices of its **Environmental Training for Sustainable Development** initiative, the Management and Training Services Division of the **Commonwealth Secretariat** supports short- and long-term training, internships and institution development for environmental policy makers, environmental 'operatives', and environmental information professionals in the Commonwealth, in various areas of the environment including biodiversity and gender. Funding support for training, institution development and publications under the aegis of the Management and Training Services Division is provided by the Fund for Technical Co-operation (CFTC).



Published by

Commonwealth Secretariat

ISBN

0-85092-548-7

Copyright

© 1998 World Conservation Monitoring Centre

Reproduction of this publication for educational or other non-commercial purposes is authorised without prior permission from the copyright holders, provided the source is acknowledged.

Reproduction for resale or other commercial purpose is prohibited without the prior written permission of the copyright holders.

The views expressed in this book do not necessarily reflect those of WCMC or its collaborators.

The designations of geographical entities in this report and the presentation of the material do not imply the expression of any opinion whatsoever on the part of WCMC, the Commonwealth Secretariat, the Darwin Initiative for the Survival of Species, or other participating organisations concerning the legal status of any country, territory, or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Citation

World Conservation Monitoring Centre. 1998. *WCMC Handbooks on Biodiversity Information Management. Volume 5: Data Custodianship and Access*. Reynolds, J.H. (Series Editor). Commonwealth Secretariat, London. ix + 24pp.

Typeset by

Bookcraft Ltd, Stroud, Gloucestershire, England

Cover design

Michael Edwards

Photography by

J.S. Donaldson (Wood's cycad, *Encephalartos woodii*); D. & I. Gordon (Mali landscape; Plant study, Ghana; Thai forest; Rock hyrax, *Procavia capensis*); IUCN/J. McEachern (Diver and fish); WCMC (Ecoregion and Africa maps; GIS work; Workshop facilitation).

Available from

IUCN Publications Services Unit
219 Huntingdon Road, Cambridge, CB3 0DL, UK
Tel: +44 1223 277894; Fax: +44 1223 277175
Email: iucn-psu@wcmc.org.uk
<http://www.iucn.org/bookstore/>

Contents

ACKNOWLEDGEMENTS. v

BACKGROUND vii

1 INTRODUCTION 1

2 BASIC PRINCIPLES 3

3 FUNCTIONS OF A CUSTODIAN 5

 3.1 Responsibilities and Rights 5

 3.2 Custodians and Owners 5

 3.3 Custodians and Users 7

4 MANAGING CUSTODIANSHIP 9

 4.1 Assigning Custodianship 9

 4.2 Managing Custodianship 11

 4.3 Reviewing Custodianship 12

 4.4 Setting Priorities. 13

5 CATALOGUES AND METADATABASES 14

6 DATA ACCESS AGREEMENTS. 15

 6.1 Overview 15

 6.2 Roles of the Hub, Custodians and Users 18

 6.3 Cost Recovery. 19

7 CASE STUDY: COMMONWEALTH CUSTODIANSHIP
 GUIDELINES, AUSTRALIA 21

8 REFERENCES 24

ACKNOWLEDGEMENTS

The generous support of the *United Kingdom's Darwin Initiative for the Survival of Species* has provided for the development of a comprehensive programme of training in biodiversity information management. This programme comprises an international training team, drawing on expertise from collaborating organisations around the world; the preparation of a training resource in the form of a handbook series and related materials; and the development of computer-based demonstration tools. Training is being promoted through the delivery of post-graduate modules, and through regional and national workshops which have received additional support from The British Council, British Airways Assisting Conservation Scheme, and contributions from participating organisations. The programme has been appropriately titled *Darwin Initiative Training in Biodiversity Information Management*.

Development of the handbooks has also benefited from experiences gained through the Biodiversity Data Management (BDM) Project, administered by the United Nations Environment Programme (UNEP) and funded by the Global Environment Facility (GEF), and related initiatives supported through the European Union (EU) and European Environment Agency (EEA). Indeed, Volume 6 draws extensively on one of the key outputs of the BDM Project, the *Guide to National Institutional Survey* (UNEP/WCMC 1998), developed in consultation with participating countries, the BDM Advisory Committee and the UNEP management team. The concept of an information cycle was developed in collaboration with the International Institute for Environment and Development (IIED) with support from the UK Department for International Development (DFID). The handbooks have been published through the generous support of the Commonwealth Secretariat.

Fundamental to the development of this programme have been the partnerships established with training organisations around the world. These organisations have worked collaboratively in hosting workshops, in reviewing the handbook materials, and in providing guidance on how regional and national training needs can be met most effectively. The training programme has significantly benefited from the input of numerous individuals working in the field of biodiversity information management. Among these individuals, particular mention goes to Professor Ian Crain and Gwynneth Martin of the Orbis Institute, Ottawa, Claire Appleby, an independent consultant, and to Drs Jake Reynolds and John Busby of WCMC for their insightful work in developing the handbook series. Thanks are also extended to Laura Battlebury for her tireless administrative and logistical support. The series

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

Many organisations are beginning to realise that knowledge is one of their greatest assets. They are also discovering that the costs of maintaining this asset, and exploiting the opportunities it creates, are significant. Given these costs, which may be significant, organisations are looking to gain maximum value from their data holdings. Increasingly, they are using data for multiple purposes and are recognising the need to develop documentation and ensure compliance with established standards. They are also beginning to understand the benefits of sharing data and collaborating with others within information networks (see Volume 4).

The development of policy-relevant information on environmental concerns often requires access to a wide variety of data sources, from numerous organisations and disciplines (see Volume 3). If the process of information production is to be efficient and cost-effective, such sources need to be readily accessible, as should the people and tools necessary to convert them into information for decision-making.

Preferred sources of data are those organisations (occasionally individuals) which are in the **best position to ensure the quality and accessibility of their datasets**, and to advise on appropriate uses. These are referred to as **custodians**.¹ If custodianship is not assigned and managed carefully, then users may face a bewildering set of incomplete and incompatible datasets, with inadequate documentation and poorly defined and inconsistent access procedures. As a result, they may find it virtually impossible, within the limited time available, to integrate data into information that will usefully support decision-making.

Custodianship is the means by which responsibility for the management of a dataset (or part thereof) is assigned to and accepted by the most appropriate organisation. Its principal aims are as follows:

- To minimise duplication of effort.
- To ensure that data are available for use (i.e. they exist and are accessible).

¹ Although most references in this handbook concern the custodianship of data, the term applies also to *tools, applications* and other *technologies* which transform data into information, or communicate that information to users (see Janzen, 1993).

- To ensure that data are quality-assured (i.e. they are valid, maintained, documented and secure).

Custodianship provides a mechanism to ensure that important datasets exist, are maintained and are accessible to legitimate users. It **ensures accountability for and reliability of datasets** within a specific jurisdiction (e.g. a sector, discipline or theme), thus ensuring that information products used by governments and other decision-makers are accurate, complete, identifiable and auditable. In summary, custodianship is the core of an efficient, responsive information infrastructure, capable of serving the interests of individual organisations or networks.

2 BASIC PRINCIPLES

As with other complex issues relating to the management of scientific information, effective progress requires the underlying principles of data custodianship to be understood and widely accepted. These principles, which are presented below, can then guide operational developments:

- Data should be managed by the organisation in the best position to do so.
- Data should be managed cost-effectively by staff who understand the data: what it purports to represent, what its characteristics are, how it was collected, what quality-assurance procedures have been applied and its limitations.
- Data should not be duplicated or fragmented in different places.
- Data should be widely available to those that have a need to make better informed decisions.

The above principles are not meant to be interpreted as rigid rules. There may be sound operational reasons why, for example, a copy of a dataset may be duplicated in another place, such as in a network hub (see Volume 4) for purposes of distribution. This may be done because the custodian is not accessible on-line, which could impede access to the original dataset. However, the risk (and cost) of violating custodianship principles needs to be understood. In the above example, the risk is that the distributable copy will become out of date, while the cost implications are that the hub will have to monitor developments in the original dataset and periodically obtain an updated copy.

The key is to manage data in such a way that they can be converted into a variety of information products, for a variety of users, thus ensuring that they are flexible enough to respond to the demands of decision-making. Ideally, every dataset has a recognised custodian although, if priorities need to be made, the custodians of **essential datasets** should be determined first. These datasets are sufficiently vital to the day-to-day operation of an organisation or network that they justify the effort and expense incurred in their collection, storage and quality-assurance (see Volume 3).

Essential datasets underpin the development of multiple information products, for multiple users, and are, thus, permanent, or at least of a lasting nature. Non-essential datasets, on the other hand, are produced by *ad hoc*, undocumented processes for quick results, and may be transient or of uncertain quality (after Janzen 1993, 1995). Unfortunately, many organisations manage essential datasets as if they were non-essential datasets.

3 FUNCTIONS OF A CUSTODIAN

3.1 Responsibilities and Rights

As outlined in Volume 4, custodianship of a dataset carries with it certain **responsibilities** (see Box 1). Each of these contributes to the well-being of a dataset and, thus, to internal productivity gains within the organisation as well as an increased capacity to collaborate with others. Naturally, custodians may harbour concerns at the prospect of providing access to their data. For this reason they are also invested with certain **rights**, consistent with broader government, corporate and other applicable policies and agreements, which determine the conditions under which the dataset can be used (see Box 1). Such rights are not intended to prevent legitimate use of a dataset. Indeed, the aim is to foster an environment in which data access is straightforward and encouraged.

Where a custodian does not have sufficient resources to undertake all of its responsibilities, certain of these may be delegated or contracted to other organisations, known as **stewards** (Janzen 1993). Custodianship, however, remains with those responsible for the content of the dataset.

3.2 Custodians and Owners

Copyright does not protect facts, so it is not clear whether ‘data ownership’ has any legal status. However, the concept is useful when describing those individuals or organisations that have some claimed intellectual property rights (whether enforceable or not) over certain data. In the majority of cases these ‘rights’ are claimed on the grounds of original collection of the data.

Although, in most cases, data owners are also the custodians, it is important to recognise that **data custodianship differs from data ownership**. Custodianship does not necessarily signify ownership, although the distinction is important only when the data custodian is a different entity from the owner. Box 2 presents a variety of situations in which this occurs.

Box 1 Responsibilities and rights of custodians

Responsibilities

- To build a dataset (with partners, as appropriate).
- To maintain a dataset (i.e. keep it up to date, abreast of standards, structured as necessary).
- To ensure the quality of a dataset (i.e. ensure that it is valid, maintained, documented, secure).
- To provide access to a dataset (to legitimate users).
- To provide advice on appropriate uses of a dataset (e.g. suggested/unwise/improper uses).
- To coordinate the development of a dataset (with appropriate partners).

Rights

- To regulate access to a dataset (depending on category of user).
- To safeguard intellectual property (e.g. acknowledgement, regulation of copying).
- To recover costs (e.g. recover market value, investment, cost of supply).

The owner retains intellectual property rights over the data, although these rights can, in practice, be virtually non-existent, as is the case with public-domain data. The owner may choose to delegate some of these rights to a custodian. The custodian may be likened to a trustee in terms of its relationship with the data. The degree of freedom that the custodian has to either use or distribute the data depends on how rights are delegated by the owner. In general, there will be a formal agreement between the owner and the custodian which specifies what the custodian is allowed to do with the data and the circumstances in which the owner needs to be consulted.

Box 2 Where data ownership and custodianship differ

- In most government jurisdictions, data may be ‘owned’ by the supreme executive authority, whereas the custodian may be just one of many agencies created by that authority, designated to act on its behalf.
- Data may be in the public domain but managed by a custodian organisation acting, at least to some extent, ‘in the public interest’.
- The owner of a dataset may be one or more persons, for example scientists or naturalists, who originally collected the data, and who maintain ownership rights, but who may be unable or unwilling to manage the data effectively over the long term.
- The owner may be an organisation that built a dataset for a particular purpose but has since lost interest in managing the data. It may then delegate its custodial responsibilities to another organisation, which becomes the custodian. The original owner may still retain some residual intellectual property rights over the data, which the new custodian would be obliged to respect.

In some cases, the custodian may choose to contract some of their responsibilities to another organisation, perhaps one of its partners in a network. Provided the contracted organisation has no authority to use or distribute the data without approval by the custodian, this arrangement would not ordinarily affect the custodian’s rights or responsibilities over the data.

3.3 Custodians and Users

Properly organised custodianship is beneficial to users of both data and information. For example, confusion over where to obtain accurate data is minimised, and reliable advice on the source, currency and completeness of information products is forthcoming. In return, **users should assist custodians by providing feedback on the usefulness of data**, and by keeping them informed of their future requirements (e.g. quality-assurance requirements). This helps the custodians plan their data collection and management strategies.

Where a user collects data on behalf of a custodian, with the intention of submitting it for entry into a dataset, this should be done according to the standards and procedures established by the custodian. Users should also return any data that they have corrected or otherwise upgraded and, in turn, the custodian needs to ensure that the upgraded data is made available to subsequent users.

4 MANAGING CUSTODIANSHIP

4.1 Assigning Custodianship

Custodianship of a dataset is normally accepted by the organisation most familiar with its history, special management requirements and potential uses. Within any particular network, such organisations may be obvious to the network's partners, allowing custodianship to be confirmed, rather than negotiated, without issue. However, where several organisations claim custodianship of the same dataset, or no custodian is apparent, the network hub, through its steering committee, may decide to **commission a review**. This could be applied to specific datasets as the need arises or, more ambitiously, it could be extended into a network-wide review (see Volume 6).

It is often the case that environmental datasets are significant to a wide range of stakeholders, not just their custodians. This suggests that greater, perhaps national needs should prevail over individual feelings of data ownership, particularly in the case of essential datasets which may be depended upon for projects of national importance. Difficult decisions may have to be made in the short term to guarantee the quality and accessibility of a dataset in the long term. As with all decisions of this nature, it is imperative that they are arrived at transparently and with the **full participation of leading stakeholders**. Box 3 lists a variety of criteria which the steering committee of the hub could use to determine which organisation is the most appropriate custodian for a particular dataset (note that the criteria are not equally important).

One way to apply the criteria in Box 3 is to select those organisations thought to be most relevant in the particular context, and assign numeric values to each criteria according to its relative importance. Thus, statutory responsibility may be perceived as being the most influential factor determining custodianship in a particular network, whereas best financial position might be considered to be the least. This allows the steering committee to 'score' potential custodians according to their suitability for the role, as illustrated in Table 1 (custodians denoted by the letter C). Analyses of this kind do not provide sufficient grounds for assigning custodianship in their own right, but may serve to focus discussion.

Box 3 Potential criteria for determining custodianship

- Statutory responsibility for management of a dataset (beware of overlapping mandates!).
- Greatest operational need for a dataset (e.g. for decision-making).
- Normally first to record changes to a data item.
- Requires the highest integrity of a data item (e.g. military precision).
- Most 'competent' to manage a dataset.
- Best financial position to manage a dataset (beware short-term effects).
- Most technical or physical resources to manage a dataset.
- Confidence of users in continuing to manage and develop a dataset (e.g. committed, no 'conflict of interest').

When deciding the custodianship of essential datasets, the overriding principle is that **each dataset should have one and only one custodian**. This is a practical and effective way of ensuring that management responsibility is assigned to every dataset which is valuable to multiple organisations and users. However, some environmental datasets are not easily packaged under a single label, and overlaps in organisational jurisdiction will occur. This can be resolved by designating one organisation as the overall custodian and encouraging others to maintain specific sub-components. An example would be a protected areas agency which manages a dataset containing, amongst other entities, data on the distribution and significance of species within its estate. Whilst it is justified in managing this dataset, the list of names used to reference the species would be managed by a more specialist custodian, such as a national museum or herbarium.

Table 1 Determining custodianship

Criteria	Relative importance	C ₁	C ₂	...
Statutory responsibility	5	✓		...
Greatest operational need	3	✓		...
First to record changes	2	✓		...
Most ‘competent’	2		✓	...
Best financial position	1		✓	...
Confidence of users	4		✓	...
...
Score		10	7	...

4.2 Managing Custodianship

Responsibility for data may need to be assigned at **several levels**. At the national level, responsibility for data themes may be assigned to separate lead organisations, such as government departments or research establishments. Land infrastructure, for example, including administrative boundaries, topography, settlements, roads and rivers, might be assigned to a national department of survey and mapping, which other national-level organisations see as the natural custodian.

At the sub-national level, land infrastructure data may be managed at a higher resolution by local authorities, and be dispatched upwards to maintain the survey and mapping department’s datasets. Thus, in reality, the survey and mapping department is the hub of a land infrastructure network, with responsibility for data management devolved to a series of sub-national custodians (see Volume 4). As such, data harmonisation — the ability to integrate the various sub-national datasets — needs to be resolved, at the outset, by agreeing appropriate data standards and protocols.

Network hubs should ensure that they provide sufficient **guidance and coordination** to custodians to enable them to contribute effectively to the network’s objectives (see Volume 4). Some obvious examples of where guidance may be

provided are the agreement of standards for data collection, storage and quality-assurance, and consistent procedures for data access (see Volume 7).

A further complication is that, while a theme may have been assigned to a particular organisation, other organisations may need to develop datasets within that theme to meet their particular objectives, and for which they would become the custodian. For example, the theme *vegetation* may be allocated to a natural resource management agency, yet the defence department may need to develop a subsidiary vegetation dataset, with attributes that show which areas heavy vehicles may traverse and which they should not because of unacceptable surface damage.

When a dataset is modified, for instance by enhancement or integration with other data, then the responsibilities and rights of its original custodian become diluted. Some management process should then define a clear point at which the original intellectual property rights and custodial responsibilities will be deemed abandoned. Essentially, it is a matter for negotiation between the parties concerned as to how to balance the requirements of the organisation that created the derived dataset with the legitimate interests of the original custodian. For instance, a decision must be made as to whether the new dataset is repatriated to the original custodian or whether it should be retained by the organisation that created it (presuming that they are capable of performing their custodial functions) (CSDC 1995).

4.3 Reviewing Custodianship

In becoming a custodian, an organisation needs to consider its data management responsibilities and ensure that it is able to meet them. If an organisation cannot meet its obligations, then it may consider **relinquishing custodianship** to another organisation. As with the initial assignment of custodianship, reassignment requires active but sensitive management, especially with datasets which have been identified as being fundamental to many organisations and users.

It is a good idea to undertake periodic reviews, perhaps every two years, of:

- the appropriateness of current custodians;
- progress in building, managing and enhancing key datasets;
- dataset documentation, including catalogues, directories and metadatabases; and
- implementation of data standards (see Volume 7).

Network hubs could consider developing a series of performance indicators for custodians, so that reviews can monitor progress against established benchmarks. In turn, custodians could report progress to their network hubs on a regular (e.g. annual) basis.

4.4 Setting Priorities

Resource constraints invariably mean that **strict priorities for data development are needed**. For this reason, it is important to identify appropriate custodians for essential datasets, and support these where specific investments in institutional capacity are required (see Volume 6). Every country, for example, needs an accurate and stable dataset representing its national boundary. This dataset must be maintained at a resolution and accuracy suitable for all major organisations and programmes, which may require the boundary in a range of scales and projections. One solution is to assign a single organisation, normally a national mapping agency, custodianship for the entirety of this dataset in its various forms.

Custodians generally build datasets for their own corporate objectives, rather than for the wider benefit of the networks in which they operate. Thus, ideally, when building datasets, organisations should take into account the needs of their fellow partners, in order to **increase the range of purposes to which the datasets can eventually be applied**. Where partners require datasets that are of finer resolution, more elaborate or, in general, are of a higher quality than that required for the custodian's own purposes, then agreement needs to be reached on how to cover any additional costs which may be involved.

5 CATALOGUES AND METADATABASES

Custodians are normally expected to document their datasets and to provide summary descriptions to potential users as required. It is generally considered to be **good professional practice** to document datasets, and custodians should aim to undertake this as a matter of routine (see Volume 7). Typical features to document include the theme, scale, completeness, currency, reliability, precision and pricing strategy of the dataset, plus details of how it was collected, its intended purposes, and the data standards and quality-assurance procedures which have been applied.

Within an information network, custodians may be encouraged to submit details to the hub on any datasets which are, at least potentially, available for use by other organisations. The catalogues which result — known as **metadatabases** in their computerised form — may be published and disseminated widely to assist users in locating the data and information they require. Due to the potentially large number of datasets available within a network, catalogues usually contain only a summary of the dataset's purpose and quality, plus the contact details of the custodian and advice on access procedures, including any costs involved.

6 DATA ACCESS AGREEMENTS

6.1 Overview

Issues relating to data access are some of the most challenging, yet important, for an information network to manage. The challenge is rooted in the legitimate concerns of both custodians and users, as summarised in Box 4. These concerns are held by individuals, project groups, organisations and governments, and cover a wide range of issues. Some of these are genuinely important,² but others are largely spurious or secondary to other issues, for example mistrust of user intentions or lack of understanding of what the user requires.

Unless fully addressed, such concerns have the potential to hinder data access and, hence, reduce interest in cooperation. One way forward is to accompany transactions with formal agreements between parties, offering tangible assurances that the concerns of both sides will be met. A key principle is that data should be made accessible **through the custodian**, or their nominated distribution outlet, not from a secondary source.

Data access agreements need not be viewed as a defensive measure intended to limit access to data. Rather, they are a **positive means of increasing trust and fair dealing** amongst the network's or other partners. As confidence grows and data begin to be mobilised more easily, the need for formal agreements diminishes until, ultimately, they may no longer be required.

Data access agreements are perceived to be difficult to negotiate, but this need not be the case. A useful step is for the network hub to draft a **generic agreement** for distribution and adaptation by custodians. In the interests of simplicity, agreements should assume goodwill on the part of users, not bad faith. Similarly, administrative and cost impediments to data access should be kept to an absolute minimum.

2 Custodians need to ensure that data which are genuinely sensitive for reasons of privacy, confidentiality or security, are adequately protected. An example would be detailed descriptions of the locations of threatened species that are at risk from exploitation.

Custodians are responsible for advising users on the potential uses of a dataset. They may also recommend specific permitted uses or, conversely, excluded uses of a dataset. This may occur when, for example, the custodian believes the data are unsuitable for certain purposes due to uncertainties or ambiguities in the dataset. It also provides a means for custodians to safeguard their intellectual property. Varying conditions may be applied to different classes of user (e.g. government, NGO, research, commercial). Typical provisions to consider when drafting a data access agreement are presented in Box 5.

Box 4 Typical concerns over data access

Custodians

- Will the dataset be misused?
- Will intellectual property be respected?
- Will the cost of supply be recovered?
- Who will be liable in the event of a problem?
- Will professional credibility suffer from the release of the dataset?
- Could confidentiality be breached by the release of the dataset?
- Does the transaction comply with internal network guidelines?

Users

- Is the dataset fit for its intended use?
- Will it be available at the right time?
- Will onerous conditions be imposed on its use?
- Is it available in a form which can be easily handled?
- What, if anything, will it cost?

In the event that incorrect data are provided by a custodian, or they are used for an inappropriate purpose, liability could fall on the owner of the data, the custodian, a third party which has provided the data, or all of these. The situation is most serious when ‘negligence’ is detected, for instance when it is established that data were poorly maintained or falsely documented. Exposure to liability is an emerging issue. The most likely grounds for liability are where dataset documentation is misleading, deceptive or negligent, or where there has been a violation of the provisions of a data access agreement (see Onsrud 1989 for a discussion).

Box 5 Typical provisions in a data access agreement

- Whether data are available for single or multiple uses, and what those uses are.
- Whether data can be retained by the user following its designated use (risking independent updating or duplication), or whether they should be destroyed (incurring subsequent costs when the data are needed again).
- Whether the data are available for non-profit or profit-generating use and, if the latter, whether any royalties would flow back to the custodian.
- Whether users may give the data to third parties (the latter would normally be referred back to the custodian unless prior consent was arranged).
- Whether copies of any publications, products and other outputs derived (even in part) from the data should be sent to the custodian.
- What form of acknowledgement the user should employ (in cases where data have been interpreted several times before use, a long list of acknowledgements — known as an ‘audit trail’ — may be necessary).
- Some form of disclaimer that protects the custodian from legal liability in the event that their data prove to be unreliable or are used for an inappropriate purpose (legal advice should be obtained).

6.2 Roles of the Hub, Custodians and Users

The primary role of the hub is to enable the network's partners to realise the benefits of cooperation. In terms of data access, these benefits comprise value to the user through access to essential data; value to the custodian for being of service (paving the way for future exchanges and access to value-added products); and value to the network for having enabled the transaction.

The hub can help to realise these benefits by facilitating the negotiation and management of data access agreements. This could be achieved through the preparation of a generic agreement embodying the principles of cooperation it wishes to invoke, i.e. that **environmental data are available to all stakeholders**, for non-commercial purposes, at minimum cost and with minimum administrative or other impediments to access and use, whilst protecting the legitimate interests of custodians.

In cases where access to an important dataset is not being provided satisfactorily, the hub may wish to scrutinise the restrictions imposed by the custodian and recommend new operating procedures. Experience has shown that many restrictions imposed by custodians are unnecessary, even arbitrary, and can impose severe constraints on data access.

The hub's advice will need to be consistent with government policies relating to information access and exchange. Where current policies are inconsistent with the principle of wide information availability, with environmental decision-making likely to be less effective as a consequence, network hubs have a role in advising the government accordingly.

Custodians are responsible for developing policies on data access which are consistent with the broader policy frameworks in which they operate. They need to ensure that the provisions in their data access agreements are **necessary to protect their legitimate interests**, and do not have the unintended consequence of inhibiting reasonable use of the data by other users.

Users are obliged to **comply with conditions** prescribed by custodians. This is crucial if the spirit of trust, so necessary to the successful operation of a network, is to be built and maintained. Users may also provide feedback to custodians in the form of advice of any errors or deficiencies encountered in the data, and an indication of their future requirements, enabling custodians to continually improve their service.

Finally, there are two responsibilities which users would ordinarily uphold even if not specified by agreement. Where practicable, copies of any information products derived fully or in part from a dataset are provided to the custodian; and the source(s) of the data used should be **fulsomely acknowledged**.

6.3 Cost Recovery

Very few organisations have a corporate objective to collect data merely to sell data. Data-rich organisations are generally in the business of selling the **value they add** to data. In other words, the collection and management of data is a means to a larger corporate end. As such, the cost of building and managing data is a cost of doing business or, more simply, a business overhead. Certainly these costs need to be recovered by selling products and services to clients, but not necessarily by charging total costs to external data users.

When raw data are provided to a user, without any significant added value, then the question arises as to what proportion of the data management overhead to charge to that user. Normally, there is no argument that the marginal costs, to the custodian, of providing those data can be charged to the user; the issue is what is a **reasonable price** for data that already exist (particularly where collected at public expense), but which were not created for the user or in the expectation that the user would require them.

Unfortunately, the development of information networks is sometimes inhibited by market-driven ideologies which authorise the cost recovery in data transactions. The consequence of this is that organisations which could contribute to the resolution of important environmental concerns are unable to do so because they lack the resources to purchase the relevant data. The resources available to environmental organisations, particularly NGOs, seldom reflect the significance to the community of the issues at stake. This is largely due to a failure of the market to internalise environmental costs and benefits.

Within a network, **cost barriers should be kept as low as practicable** to facilitate data exchange. However, depending on the financial positions of the organisation concerned, and the types of user who require access to the data, different strategies for cost recovery may be applied. Not all of these are monetary-based, as illustrated in Box 6.

One way of facilitating data access in hostile policy environments (e.g. over-bureaucratic or market-driven) is to develop datasets in partnership with other organisations, or to barter data and other services (see Volume 4). There is no reason why such arrangements should not include commercial partners.

Box 6 Strategies for cost recovery

- Supply data totally free.
- Supply data free to reciprocating users (i.e. barter).
- Recover immediate cost of supply (i.e. time, consumables, energy etc. to process the request).
- Recover incremental costs (i.e. of managing the data and combating depreciation).
- Recover development costs (i.e. the initial investment in building the dataset).
- Recover 'market value' (i.e. the maximum cost which the user will pay).

7 CASE STUDY: COMMONWEALTH CUSTODIANSHIP GUIDELINES, AUSTRALIA

Australia has formed the Commonwealth Spatial Data Committee (CSDC) as the peak coordinating body (network hub) for spatial data management at the national level. It consists of senior governmental representatives. One of its tasks has been to draft the *Commonwealth Custodianship Guidelines setting out the Rights and Responsibilities of Spatial Data Custodians and Lead Agencies* (<http://www.auslig.gov.au/pipc/csdc/csdcguid.htm>), the executive summary of which notes:

“The principle of custodianship should be applied . . . as a matter of good practice, convention and convenience. It is simply the only way that data can be managed in an orderly fashion . . .

The identification of custodians and the effective operation of custodianship are necessary for the successful management of government spatial data...

To help improve [the management of . . . spatial data] the Committee will implement a system of lead agencies and custodians of Commonwealth spatial data, and outline their rights and responsibilities . . .

These guidelines are designed to assist lead agencies and data custodians in developing improved practice in spatial data management . . .

A lead agency will be responsible for government-wide coordination of a data class or category. This includes development of standards, and coordination of data acquisition so as to avoid duplication. The CSDC will provide policy guidance and oversee the lead agency system. Lead agencies will be identified for broad categories of data after negotiation within CSDC . . .

CSDC will maintain a register of spatial data custodians. The Register will define the custodian agency or agencies and data category, data items, geographic coverage and any other feature necessary to define the dataset . . .”

The *Guidelines* discuss the distinction between lead agencies and custodians, noting that the responsibilities of a lead agency are primarily those of policy-development and coordination, and detailing a set of specific objectives for cooperation. Clearly, in this example, lead agencies correspond to network hubs for the thematic areas in

which they have expertise. The CSDC itself corresponds to an over-arching hub seeking to coordinate the efforts of lead agencies.

The custodian of a dataset is described as ‘the nominated body, or person responsible for the development and/or the management of that dataset, and who has the right to determine the conditions on which those data may be used or released’. The *Guidelines* stress that:

“All spatial data collected . . . forms part of the . . . corporate spatial data resource. Individual agencies involved in the management of the spatial data act as custodians on behalf of the Commonwealth. The Commonwealth Government delegates operational responsibility for the data, and for implementing Commonwealth policy with regard to it, but retains the ultimate responsibility for the data.”

Custodianship is not necessarily synonymous with ownership of either the data nor the copyright of that data.”

Among the benefits of custodianship, the *Guidelines* note that improved coordination assists in the avoidance of duplication of effort and the unnecessary costs which this causes, thus achieving greater benefits from the investments in data development which are made. The assignment of custodianship is recognised as being a complex process, requiring formal negotiation and documentation of the outcome. A list of essential datasets, together with suggested lead agencies and possible custodians is located at <http://www.auslig.gov.au/pipc/csdccust.htm>.

Lead agencies are highlighted as a mechanism to facilitate higher-level coordination of data themes, including developing broad initiatives, future projects and agreement of standards. Each identified lead agency is responsible for disseminating information about its category of data, particularly to potential users, and for promoting access to the data. This includes seeking to remove impediments to and improving the efficiency of data transactions.

A major role of the lead agency is to avoid duplication of effort in data collection, by providing avenues for communication between the network’s partners and mechanisms to identify and coordinate data collection activities. However, the lead agency is not empowered to determine priorities for data acquisition; that is the right of the custodians.

The *Guidelines* further note that standards for recording and managing data need to be considered, both by lead agencies and custodians. In particular, standard methods for selecting features and attributes, for assigning meanings and values, for ensuring data quality, and for regulating data transfer are noted. Indeed, it is recommended that custodians seek compliance with established quality management standards in order to ensure the quality of their datasets.

In this example from Australia, custodians are entitled to levy a charge for access to their data. The amount charged depends on several factors, including Government charging directives, the costs of collecting, storing and distributing the data, market rates, the intended uses of the data, and the conditions of any pre-negotiated arrangements.

More information about the Commonwealth Spatial Data Committee (CSDC) can be obtained from <http://www.auslig.gov.au/pipc/csd/csdmain.htm>. Details of the policy governing data transactions is available at <http://www.auslig.gov.au/pipc/csd/csdcsdti.htm>.

8 REFERENCES

- Janzen, J. 1993. *Management Guide to Custodianship*. Ministry of Forests, BC, Canada. (<http://mofwww.for.gov.bc.ca/isb/datadmin/s35.htm>)
- Janzen, J. 1995. Presentation Summary: *Corporate Data Categories*. Corporate Spatial Database Meeting, November 20/21, 1995, Ministry of Forests, BC, Canada. (<http://mofwww.for.gov.bc.ca/isb/datadmin/summpres.htm>)
- Onsrud, H.J. 1989. Legal and Liability Issues in Publicly Accessible Land Information Systems, *Proc. GIS/LIS*, Vol. 1. pp 295-300.

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>