BIODIVERSITY DATA MANAGEMENT
(Document 1)

DATA FLOW MODEL

in the context of the
Convention on Biological Diversity

**WORLD CONSERVATION
MONITORING CENTRE**

The mission of the
World Conservation Monitoring Centre is to provide
information on the status, security and
management of the Earth's biological diversity.

BIODIVERSITY DATA MANAGEMENT
(Document 1)

DATA FLOW MODEL

in the context of the
Convention on Biological Diversity

United Nations Environment Programme

March 1995

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF ANNEXES

1 INTRODUCTION

1.1 Background

The *Convention on Biological Diversity* (CBD) was signed at the United Nations Conference on Environment and Development in Rio de Janeiro in June 1992 by 154 nations and subsequently came into force in November 1993. Article 7 of the Convention is concerned with identification and monitoring activities to support Articles 8 to 10 (*in-situ* conservation, *ex-situ* conservation and sustainable use of components of biological diversity). Contracting parties are required to identify components of biological diversity important for its conservation and sustainable use (Article 7a); to identify activities likely to have adverse impacts (Article 7c); and to monitor the status of both components and threats (Articles 7b and 7c). Specifically Article 7d identifies the requirement to:

> *"Maintain and organise, by any mechanism, data derived from identification and monitoring activities".*

Having recognised this clearly identified need for management of data in support of national planning related to biodiversity, the United Nations Environment Program (UNEP), in collaboration with the World Conservation Monitoring Centre (WCMC), designed and submitted to the Global Environment Facility (GEF), a project proposal entitled *Biodiversity Data Management Capacitation in Developing Countries and Networking Biodiversity Information (BDM)*. This proposal was endorsed and subsequently a sub-project was established between UNEP and WCMC for *Development of Supporting Materials for Biodiversity Data Management and Exchange*.

The sub-project has produced an interlinked package of resource materials to assist in national capacity building. There are four principal components of this package:

Document 1. Data Flow Model
  (This Document)

Document 2. Guidelines for a National Institutional Survey
  - to provide guidance to countries in conducting a survey and assessment of the capacity of existing national institutions to support biodiversity information management.

Document 3. Guidelines for Information Management
  - to facilitate the development of capacity for information management and exchange as required by the CBD.

Document 4. Resource Inventory
  - the core output of the project; a collection of reference directories, guidelines, and standards relating to biodiversity information management.

The Data Flow Model is intended to identify in a formal structure the relationships between components of biodiversity data, from acquisition through to use in national strategy development, planning, and monitoring for implementation of the CBD.

1.2   Information Requirements of the CBD

The CBD has three main purposes:

- the conservation of biodiversity
- encouraging the sustainable use of biodiversity
- the equitable sharing of the benefits of the use of biodiversity.

The information required by a country to meet these objectives is wide ranging, going beyond the normal boundaries of "conservation" or "environmental" information. A number of Articles of the CBD require or imply the need for facilities for the management and exchange of biodiversity information. Articles 7d, 12c, 13b, 15(7) and 16, each identify information management and exchange requirements, and Article 17 explicitly indicates *"access to and transfer of technology among Contracting Parties are essential elements for the attainment of the objectives of this Convention"*. A clause-by-clause analysis of the implied information requirements is given in Annex 1, and summarised below.

The CBD identifies three main categories of biodiversity information:

- ecosystems and habitats
- species and communities
- described genomes and genes of social, scientific or economic importance.

To this basic list one must add:

- the scientific and technical information required to measure, assess and take decisions on appropriate action

- bio-technology, its value and risks

- local knowledge of traditional uses and values of biological resources

- interrelationships between biodiversity, human actions, laws and conditions, economics and development.

The Report of the Open-ended Intergovernmental Meeting of Scientific Experts on Biological Diversity - Second Session (UNEP/CBD/IC/2/11) provides in its annexes further indication of the scope of the information and technology of biodiversity. For instance, Annex II lists six major categories of technology *"relevant to the identification, characterisation and monitoring of ecosystems, species and genetic resources"*:

- classification technologies for terrestrial, marine and other aquatic ecosystems

- ecosystem evaluation technologies

- biogeographical mapping technologies

- isolation, characterisation and classification technologies (for terrestrial, marine and other aquatic organisms, for plants, animals microbes and genes, and for indigenous and non-indigenous organisms)

- technologies to determine species and genetic resource status

- key enabling technologies including, information technology, advanced biochemical and molecular technologies, risk assessment etc).

These main headings were further subdivided into a large number of classes from "biogeography", "ecosystem function" and "traditional knowledge" through to "abundance, distribution and range" (of species) and "biotechnology".

The following paragraphs, adapted from the United Nations Environment Programme (UNEP) *Guidelines for Country Studies on Biological Diversity* (1993), identify some of the types of information which might be important, particularly for the initial assessment and strategy development.

Biological
This is the primary focus of biodiversity conservation - the core data which includes the requirements for species, ecosystems and genetic resources, covering issues ranging from status and distribution of resources to functional relationships and the development of tools to support the science.

Physical
Information on physical factors such as climate, topography and hydrology allows biological data to be placed within a physical context, and also allows for the development of predictive models (as the distribution of many species and vegetation types can be predicted by a combination of physical characteristics). Physical factors can also have a significant effect on potential use of resources, and on management options.

Socio-economic
The use and abuse of biological resources is essentially a function of socio-economic factors. Important data might range from monitoring of forestry or fisheries practice, to the impact of farming methods, or the distribution of population centres and transport routes. Perhaps as significant is accessibility to natural resources, and the uses that local peoples make of these resources. The latter often form an essential, but perhaps invisible part of the local economy.

Costs and Benefits
In order for management of biodiversity to be efficient, it is necessary to know the true value of biodiversity and the costs and benefits of management options. This needs to cover questions such as the costs of managing protected area systems, the level of income derived from tourism, and the value of indirect benefits such as watershed protection. Methods for assessing some of these values are only just being developed, and further dissemination of information on the methodologies will be required.

Pressures and Threats
Identifying and monitoring both potential and actual threats to biological diversity is an important component of any information management programme aimed at improved management of biological resources. However, the latter may need to look beyond immediate physical causes and effects, to the underlying impact of human activities (which links threats to socio-economic factors).

Sustainable Management
Conservation of biodiversity is about effective and sustainable resource management. To assess that management, information will not only be required on the biodiversity itself, its status and distribution, but also on current and past management activities, especially on the use of biological resources. For example, information is likely to be required on a range of factors concerning protected areas, plus on effective management regimes and technologies in a range of protected and unprotected habitats.

Sources and Contacts
Information is also required on information models, standards, and technologies, and on appropriate agencies and experts who can be contacted. This may include bibliographic information on who has published what, where, basic information on names and addresses of appropriately qualified experts, sources of information on reliable and appropriate models, and metadatabases.

Interrelationships
The above paragraphs begin to illustrate the extent of the interrelationships between the information that might be required in order to study and manage biodiversity more effectively. It is essential that these interrelationships are kept clearly in mind when planning information management strategies. Comprehensive forecasting of the effects of these interrelationships is also necessary for efficient information sharing.

Another method of sub-dividing the information requirements of the CBD is the following eight-point classification, which reflects the way in which national and international agencies are organised to manage biodiversity information:

Conservation
Encompassing information on species, habitats, protected areas, biodiversity indicators, wildlife, etc.

Genetic Resources
Encompassing agriculture, agricultural research, gene banks, use of genetic resources for benefit of mankind, traditional use, genetic threats, etc.

Technology
Encompassing information on the technology of biodiversity monitoring and assessment, such as data collection technology, computer systems and telecommunications, remote sensing, geographic information systems, database techniques and standards.

Bio-technology
Encompassing a forum for interchange of information on research and application of bio-technology.

Environmental Statistics/Economics
Encompassing resource utilisation, value of biodiversity, land use, industrial outputs, equitable sharing of benefits, natural resource utilisation, trade, economics, etc.

Policy
Encompassing policy development, modelling, decision support systems and technology, empowerment and public consultation techniques, etc.

Human Factors
Encompassing population, human health, social conditions, indigenous knowledge, and their relationships to biodiversity.

Environmental Law
Encompassing environmental legislation, conventions, protocols, regulation, standards, etc.

It should also be noted that Article 18(3) requires the establishment of "a clearing house mechanism to promote and facilitate technical and scientific cooperation". This clearing house mechanism is for exchange **between** countries, but it is clear that a proper response to the CBD requires information exchange, integration and assimilation **within** each country, as well as building of capacity, to effectively utilise the clearing house mechanism. This document is intended to facilitate the development of such a national biodiversity information system.

The more detailed levels of the data flow model presented in this document give emphasis to the traditional areas of biodiversity information; that is conservation, genetic resources, and environmental statistics/economics. However, later editions of the model are planned to cover the other equally important areas.

1.3   Approach to the Data Flow Model
A distinction is often drawn between "data" and "information". Data normally refer to facts which result from measurements or observations (such as wildlife counts or the chemistry of a soil sample), whereas "information" is produced by analysing and interpreting data, usually with the intent to communicate ideas and facilitate decision making. The transformation of data into information may include processing the data using statistical techniques, analysis through models, selection and abstraction, and often expert human interpretation, for instance of the significance of patterns and trends.

An "information system" is essentially a structured set of processes (and associated people and equipment) for converting data into information, **and** for presenting it in forms which are useful for communication and decision making. Often the modern information^Xxystem will utilise computers in some of the processes, and for storage, but this is by no means necessary. The principles of information management remain the same whether or not

---

computers are used - the need for data to **flow** from process to process, the need for well defined **processes** (of analysis and integration), the need to **store** and maintain data (and information), and the need to present or **output** the information in useful forms. Some or all of the processes may be manual, requiring specialised knowledge and interpretation. In considering a Data Flow Model for the CBD, no presumption is made on the extent and nature of the use of computer technology. The Data Flow Model is intended to provide an outline of the processes for information management under the Convention which is independent of the extent to which computer systems are employed or which particular hardware and software are adopted.

Strategy development and decision making in response to the CBD requires information which is integrated and summarised to a very high level, the result being many stages of processing removed from the original raw observations of the field scientist. A national information system supporting the CBD will be characterised by a series of summarising and integrating steps. At the lower levels of the process, the outputs will require analysis and interpretation by qualified specialists in sectoral institutions. However, as the information becomes more refined, policy analysts and strategists will be required.

As noted and implied in the CBD and related documents (eg Country Study Guidelines, and the reports of Expert Panels established to follow up the CBD), the range of potential information types varies widely in biodiversity, with so many other types of information necessary for its management and understanding. The information required includes numeric, categoric, spatial and textual data occurring in a variety of forms and using a mixture of different media. This illustrates the broad scope and complexity of the data to be collected, exchanged and analyzed, and the potentially complex processes required to use it effectively.

It is also clear that the data derived from national monitoring activities depends upon the specific threats identified in the country concerned. This means that no universally applicable data requirements can be determined.

Consideration of these factors (breadth of information type, and country-to-country variability) resulted in the decision to produce what is termed a "generic" data flow model. This model has been developed through analysis of the processes that contracting parties are expected to undertake to implement the CBD, and of the broad categories of data required by these processes. An overview of the processes is given in Section 2 of this document, and a further level of detail is described in Section 3. The various types of data used in these processes are discussed in Section 4 and data models are also suggested, again starting with an overview and then expanding detail in selected areas. Applications of the models are also illustrated in this section. The intent of this document is to provide, at the national level, a sound basis for information management system designs that will:

- facilitate the presentation of biodiversity information to decision makers
- have an underlying common structure
- serve the goals of the CBD.

1.4 Conceptual Overview of National Activities for the CBD

1.4.1 Activities

The overall process for the implementation of the CBD within a country is depicted in Figure 1.1.
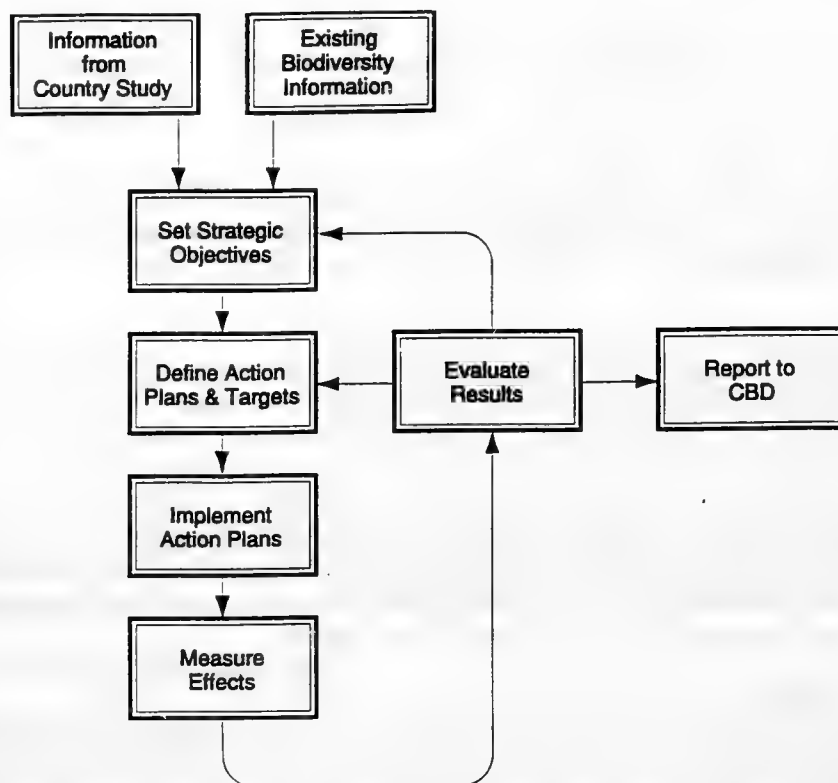
Figure 1.1: Overview of National Activities in Support of the CBD

Each component of the figure is described below.

Information from Country Study

The purpose of the *Country Study* (UNEP, 1993), recommended as the first stage in implementation of the CBD, is to provide information of various kinds to be used in the formulation of national strategies and plans for the conservation of biodiversity. The data will also provide a baseline for monitoring and assessment of the effect of measures taken (see CBD Articles 6,7).

Existing Biodiversity Information

Much information which can be used in the formulation of national strategies and action plans for a country is undoubtedly already in existence within the country, in neighbouring countries and/or with international agencies. The organisation and availability of the information will vary considerably from country to country and this will influence how easily

and effectively it can be used (see CBD Articles 6,7 and Document 2).

Set Strategic Objectives
The setting of strategic objectives must start with the identification of the components of biodiversity of importance to conservation in the country, then proceed with identification of existing and potential threats, and evaluation of the economic implications of any conservation measures. The objectives will be established at several levels of detail and jurisdiction, and they will be integrated as far as possible into relevant policies and sectors (see CBD Article 6a).

Define Action Plans and Targets
Based on the strategic objectives, priority areas for action will be defined and specific targets set. The action plans should include an indication of how progress towards the targets should be measured (see *Measure Effects* below and CBD Article 6a).

Implement Action Plans
The defined action plans should be funded and implemented. This will involve multiple organisations reflecting the different levels of strategic objectives noted above (see CBD Articles 6-11).

Measure Effects
The success or failure of actions to meet defined targets should be measured. This may require action after a specific time period, a monitoring programme on a more continuous basis, or a combination of the two. It is important that the measurement process and targets be discussed fully in the action plan (see CBD Article 7b).

Evaluate Results
The measured effects should be compared with defined targets, allowing an evaluation of the actions to be made. As a result, it may be necessary to update the action plan, or go back a step further and re-examine the strategic objectives. For example, further data may need to be collected to supplement existing biodiversity information. There should be an iterative cycle of planning, implementation, measurement, and evaluation (see CBD Articles 7b, 7d).

Report to CBD
The exact reporting requirements of the CBD have not yet been defined, but are soon to be addressed by the Secretariat. Once the requirement is fully defined, its place in the overall process will be apparent, permitting additional detail to be added to Figure 1.1 (see CBD Article 26).

Some rectangles in Figure 1.1 represent information management "processes", and others data collections or "datastores"; it is not necessarily clear which is which. In the terminology of Section 1.3, there would seem to be six processes and two datastores, although clearly, data are associated with "Evaluate Results" and "Measure Effects". The two datastores may serve as information inputs or outputs or both. The arrows joining the four boxes in some cases imply dataflow, in others some sort of action or sequence of events. This type of diagram may give a good conceptual overview, but it is not consistent in meaning. For the purpose of evolving a useful data flow model, to provide a more useful representation, it is

valuable to separate the "process" elements (what is done with the information) from the information itself. This provides for a consistency of meaning within the diagrams and independence from implementation methods (eg manual, computerised, or mixed). The method chosen and symbolism used are defined in Section 1.5 below.

1.4.2 Participants

Because of the broad scope of information required to develop effective strategies, a wide range of institutions and agencies are obliged to interact and participate in the activities depicted in Figure 1.1. The participating institutions will include those concerned with both research and policy in economic and social issues, as well as the environment and natural resources. National and sub-national agencies might include those responsible for statistics, health, education, economic development and planning, social development, science and technology, tourism, industry, land tenure and management, and law, as well as renewable and non-renewable resources, environment, museums, herbaria, national parks, heritage, and wildlife.

Participants will also include national and international NGOs, educational institutions, the corporate sector, multilateral and bi-lateral development agencies, scientific and social councils. Ideally, these all work in partnership in an atmosphere of sharing (similar to the concept of the international Clearing House Mechanism) to provide the necessary flow of information required to fulfil the objectives of the CBD. A further parallel between the CBD Clearing House Mechanism (WCMC, 1994) and national biodiversity information management process, is the concept of a linked series of specialised institutions (which may themselves be networks or "clearing houses") connected via a hub as depicted in Figure 1.2.

1.5 Methodology and Symbolism

1.5.1 Data Flow Diagrams

Following the methodology developed by Yourdon (1979), data flow diagrams are used to illustrate the flow of data between processes. These are commonly used to analyse an operation or system (eg the operation of a business) into elemental processes which are clearly understood, and to define the data required in those processes. The conventions adopted in this document are illustrated in Figure 1.3.

Operations may be expressed as a number of "processes" shown in rounded rectangles, labelled with a single digit at the first level. Each process may be broken down into (sub) processes which in turn may be split further, and so on, an extra decimal label being added at each level. Data used in the overall operation may be in one of two types of "datastore" - external, meaning generated outside the overall operation (depicted as a plain rectangle), or internal, implying that the data are generated by one of the defined processes (depicted by an open-ended rectangle).

The directional lines between processes and datastores indicate the direction of data flow. For clarity, it is conventional that each data flow diagram should contain a limited number of process boxes (4-6) and datastores.
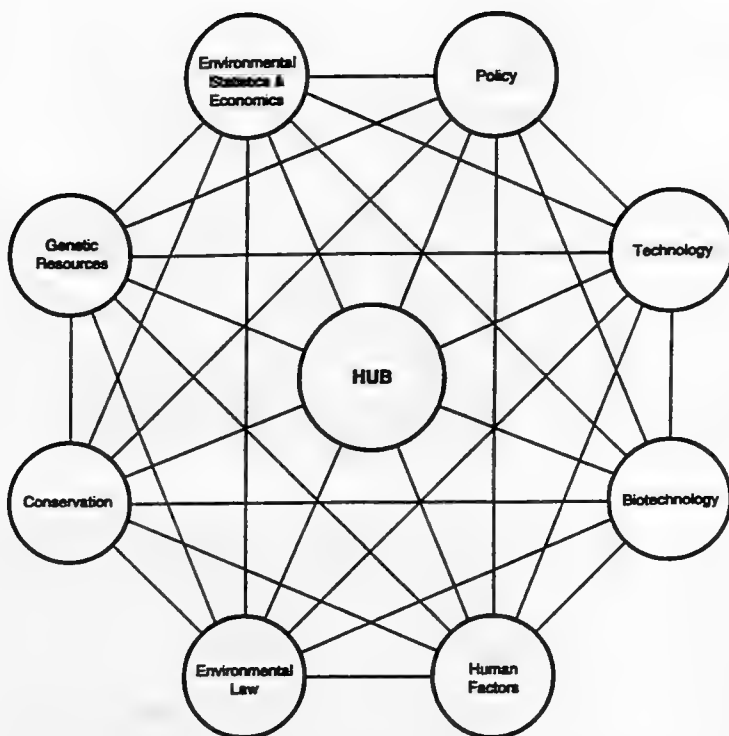

Figure 1.2: Conceptual View of a Cooperative Clearing House Network

Note that *datastore* is a generic term, and refers to any logically related collection of data. The data or information held in a datastore may not be all physically in the same institution, and may be in hard copy and/or electronic forms.

In this document, the overall operation is the implementation of the CBD within a country. The content of the datastores depicted in the data flow diagrams in Sections 2 and 3 outline the basic categories of information required. The detailed specification of the databases required to store and process these data must be determined by individual countries on the basis of their own particular needs and priorities, and the information management capabilities in place. However, given the underlying nature of the data, plus the kind of analyses and outputs frequently required, it is suggested that where computer systems are employed, the most effective solution is a relational database management system (RDBMS) linked to a Geographic Information System (GIS). The former allows extensive manipulation and reporting of non-spatial data, and the GIS extends these functions into the spatial domain, allowing data sources to be integrated and analyzed to provide outputs in a variety of forms, including graphs, tables and maps.

In the text, processes are identified in *italics*, and datastores in ***bold italics***.
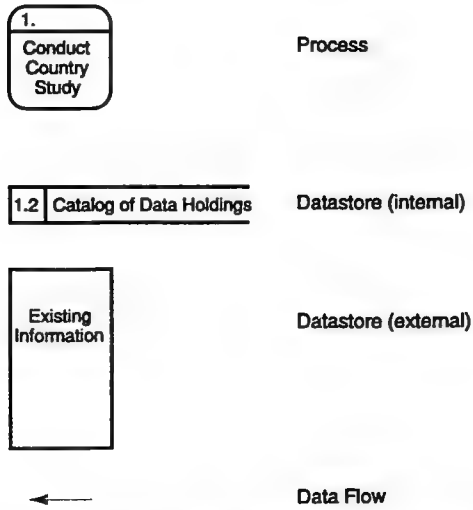
**Figure 1.3: Symbolism Used in Data Flow Diagrams
(after Yourdon 1979)**

1.5.2 Data Models

The content of the datastores is elaborated in Section 4. The discussion takes place within the context of RDBMS/GIS technologies, and methodologies associated with these are used in modelling the data. The basis is an "entity-relationship" approach (Chen, 1976), extended to encompass spatial elements.
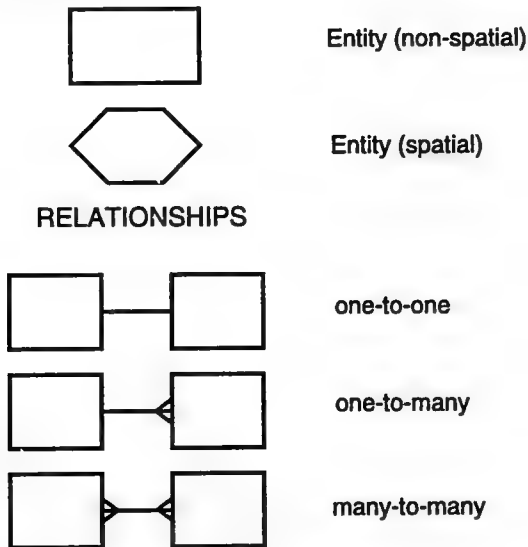
**Figure 1.4: Symbolism Used in Entity-Relationship Diagrams
(after Ashworth & Goodland, 1990)**

An "entity" is a item of interest whose attributes (properties) are being measured or recorded. For instance, an institution might be an entity with attributes of location, name, year established, mission, etc. The notation used is illustrated in Figure 1.4 and follows that of Ashworth and Goodland (1990). Thus rectangles represent non-spatial entities, lozenges represent spatial entities (points, vectors or polygons), and connecting lines show relationships between entities. The latter comprise three types as illustrated.

In the text, entities are identified in **bold.**

It should be noted that data models are to some extent subjective. Thus two individuals may produce distinct and valid models of the same data, reflecting the different objectives of their applications. The models presented in this document are generic, since the intention is to provide a framework which can be modified to meet specific situations.

2   FIRST LEVEL PROCESSES

The first level data flow diagram is depicted in Figure 2.1. This level of analysis identifies 5 basic processes and one very general datastore. Each of these is elaborated below.

INFORMATION
(Datastores)
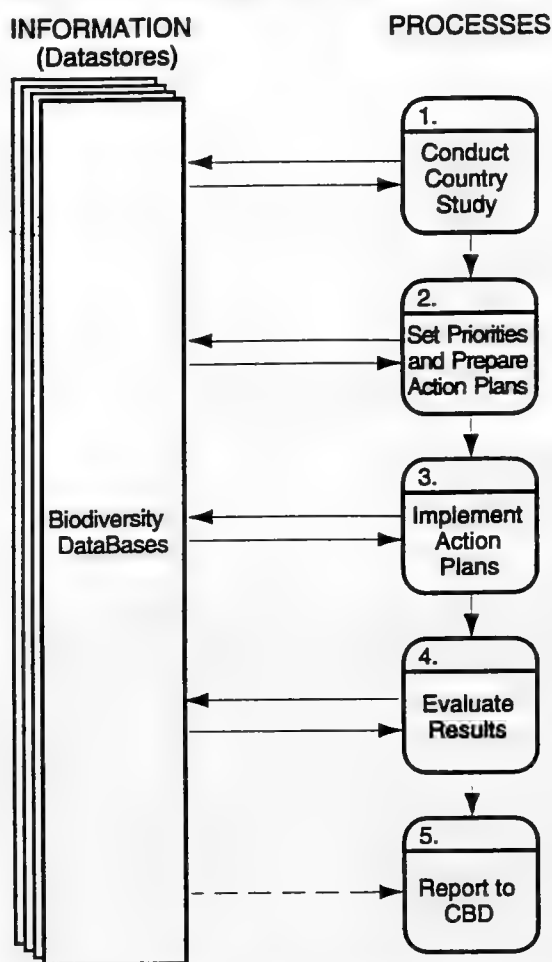
PROCESSES

Figure 2.1: CBD Data Flow Diagram, Level 1

2.1   Conduct Country Study

This process is well defined in the *Guidelines for Country Studies on Biological Diversity* (UNEP, 1993), where it is indicated that the goal of the Country Study is to initiate a process of improved biodiversity planning that will stimulate the action necessary at the national level to implement the CBD. Specific objectives include the provision of an information base for biodiversity planning and management through gathering and assessment of data required for decision making. This includes information on population, economics, environment, and so on, as well as biological datasets *per se*. Thus the *Biodiversity Databases* shown in the figure depict layers of information from several sectors to be taken into account in biodiversity management.

2.2   Set Priorities and Prepare Action Plans

This is a combination of the setting of strategic objectives and defining action plans described in Section 1.3. As noted the objectives are to be integrated with policies in relevant sectors. For example, the general objective "increase the area of natural habitat under protection" might contain "increase the area of protected forests" and/or "increase the area of protected wetlands". The action plans specify targets which will be, as far as possible, quantifiable results over specific time periods, for example "increase protected forest areas by 200 sq.km over the next two years".

2.3   Implement Action Plans

This is as described in Section 1.3, and represents the totality of all the actions taken by the institutions involved in implementing the CBD in a particular country, including measurement and additional data collection.

2.4   Evaluate Results

This is as described in Section 1.3. Note that this process includes the measurement of effects which may lead to revision of data collection plans.

2.5   Report to CBD

This is as described in Section 1.3.

2.6   Biodiversity Databases

It is unlikely that all the information required for biodiversity planning will be integrated into a single database at one site. For example, a plant species database may be maintained by the national herbarium, whereas data on protected areas may be managed by the country's national parks agency. In addition, it is clear that other information sources are required, such as baseline data for the country (eg infrastructure), physical environment data (eg soils, hydrology, geology and climate), socio-economic data (eg demographics, health, local use of resources, and land-ownership), all of which will be maintained by the agency with the relevant mandate (the custodian). The *Biodiversity Databases* shown in Figure 2.1 therefore represent the total collection of data required for all of the processes involved in the implementation of the CBD. All of the top level processes shown use the data; the processes of conducting the country study, setting priorities, implementing plans and evaluating results will add data to the overall store. In the next section, the processes are broken down into components and the very general datastore, *Biodiversity Databases*, is further sub-divided.

3 SECOND LEVEL PROCESSES

This section contains the level 2 data flow diagrams for each of the five processes outlined above (using the methodology and symbolism described in Section 1.5). In each case, the data flow diagram is followed by an expansion of the second level processes including a description of related datastores.

3.1 Conduct Country Study

The first level 2 data flow diagram is Conduct Country Study (see Figure 3.1). Each of its processes is described in subsequent paragraphs.
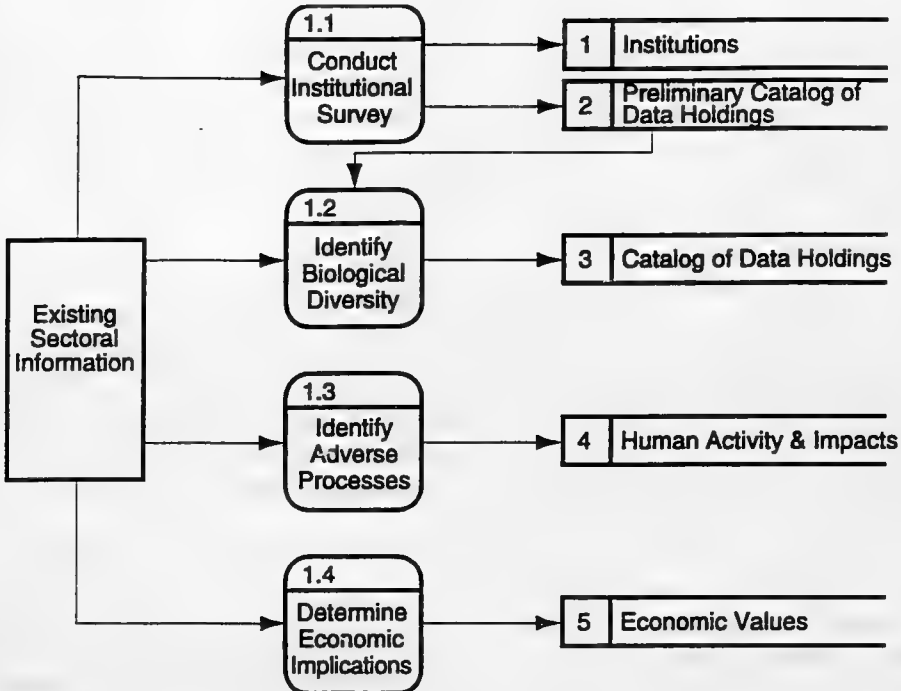
3.1.1 Conduct Institutional Survey

The *Guidelines for Country Studies on Biological Diversity* (UNEP, 1993) explicitly identify the need for an initial assessment of the country's capacity for conservation and sustainable use of biodiversity. The Guidelines suggest that the information required for this includes institutional capacities, human resource capabilities, available technological facilities and information resources in place, and that an institutional survey would be undertaken to acquire such knowledge. These suggestions are expanded in considerable detail in *Guidelines for a National Institutional Survey* (Document 2), one of the other components of this project (see 1.1). The latter include suggested methods of carrying out such a survey and details of the type of data to be collected.

The two outputs of the *Conduct Institutional Survey* process are related metadatabases. The first contains institutional information (see Section 4.3) such as staff skills and technological facilities (institutional capacity). The second catalogues the datasets (information resources) held by the institutions (see Section 4.4). This metadatabase is labelled "preliminary", since the institutional survey process involves only a cursory review of the datasets.

3.1.2 *Identify Biological Diversity*
This process ties directly to Article 7(a) of the CBD, ie the target is identification of components of biodiversity important for its conservation and sustainable use. The Convention does not include any definitive lists in this regard and items are to be defined as appropriate for the individual country. An indicative list in Annex I of the Convention (see Section 4.5) provides some insight into the nature of the required data. Clearly this process will involve examination of existing data, including those held by international agencies. The *Preliminary Catalog of Data Holdings* will provide pointers to relevant datasets in national institutions. As the content of the datasets is used in subsequent planning processes, catalog entries can be confirmed and detail added where necessary, producing the *Catalog of Data Holdings*. This datastore may be used identify significant gaps in data resources.

3.1.3 *Identify Adverse Processes*
Again this process ties directly to an article of the Convention, namely Article 7(c). Direct threats to biodiversity (ie adverse processes) include deforestation, drainage of wetlands, emission of pollutants, urbanisation and the spread of invasive introduced species. Indirect threats are less well known, but should nevertheless be considered. However, as in Section 3.1.2, these may differ greatly from one country to another. The identification of adverse processes may involve integrating and interpreting data from a wide range of institutions. The data may already suggest specific threats, or may indicate reductions in biological resources for unconfirmed reasons.

3.1.4 *Determine Economic Implications*
The process of determining economic implications is documented in Section C of the Technical Annex to the *Country Study Guidelines* (UNEP, 1993). This process includes estimating the economic value of the benefits resulting from the sustainable use of biodiversity, and quantification of the costs of current and proposed conservation actions.

3.2 Set Priorities and Prepare Action Plans
The level 2 data flow diagram for this process is illustrated in Figure 3.2.

3.2.1 *Establish Strategic Objectives*
The process of establishing strategic objectives in support of the CBD involves consultation amongst key institutions to identify the principal objectives in the context of primary threats, economic values and the capacity of institutions to support actions for biodiversity. It should make use of the four datastores resulting from the *Conduct Country Study* process, and involve analysis and interpretation of a range of existing sectoral data sources. Priorities should also identified in the process, in terms of human activities and impacts, and economic values. The output datastore will contain narrative descriptions of the strategic objectives. Attention is drawn to the paper *National Biodiversity Strategies* (UNEP/WRI, 1994).
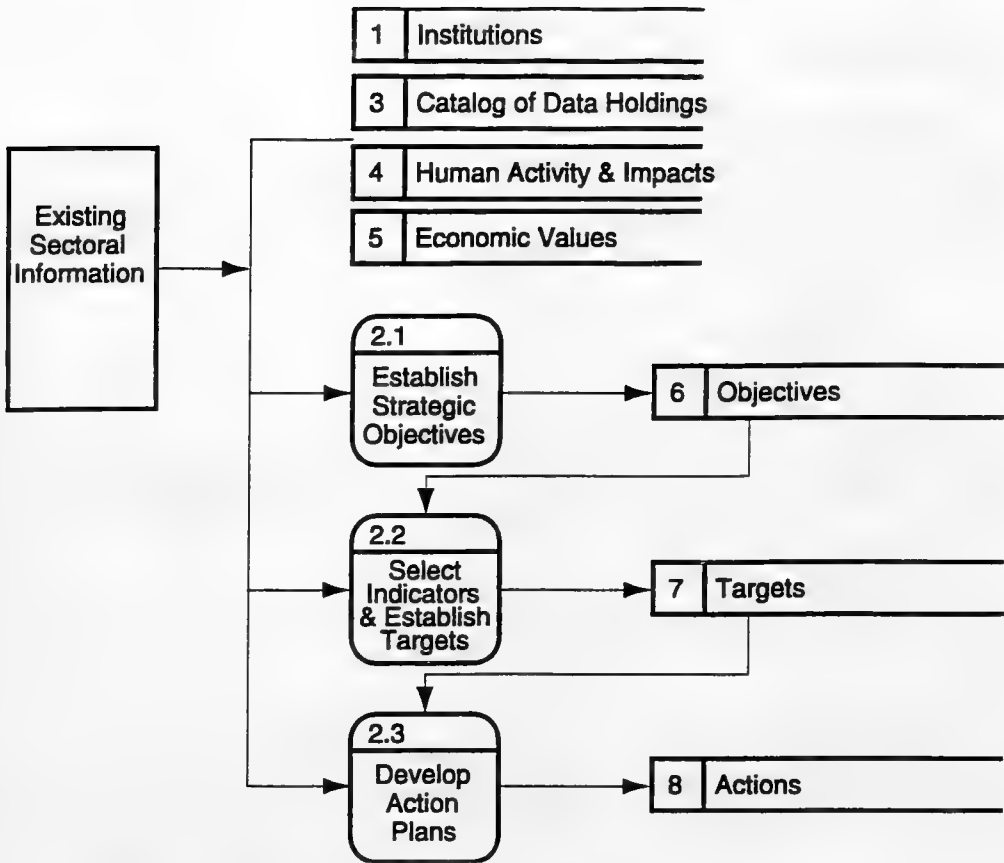
Figure 3.2: Set Priorities and Prepare Action Plans (DFD - Level 2)

3.2.2 Select Indicators and Establish Targets

In order to measure the results of actions, appropriate indicators should be chosen along with target values and critical thresholds. Indicators should be quantitative where possible, such as "to have protected areas totalling 5% of each ecosystem in the country by the year 2010". Consideration should be given to the paper *Biodiversity Indicators for Policy Makers* (WRI/IUCN, 1993). The process makes use of defined *Objectives*, as well as other inputs such as the *Catalog of Data Holdings*.

The output datastore *Targets* includes the definitions of selected indicators, methodologies for estimating them, and specific target levels. It may comprise of a mixture of textual and numeric data.

3.2.3 Develop Action Plans

The planning process must include the estimation of the costs of proposed actions, and define the institutions responsible for each task. This process therefore attempts to reconcile the *Targets* identified in Process 2.2 with available institutional capacity (see *Institutions*). The output *Actions* lists tasks, responsible institutions, required legislation and regulation, data collection and monitoring plans, and associated costs and timetables. Although the output is

mainly textual in nature, it might benefit from organisation under an automated project management system.

3.3 Implement Action Plans
The level 2 data flow diagram for this process is illustrated in Figure 3.3.
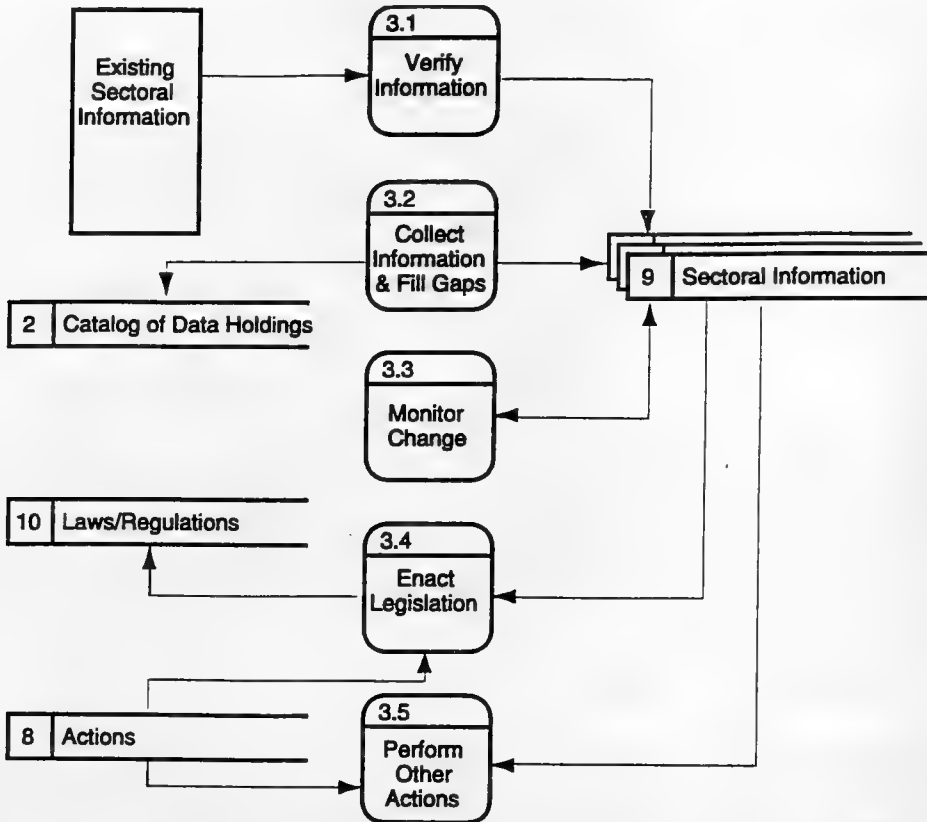
Figure 3.3: Implement Action Plans (DFD - Level 2)

3.3.1 Verify Information
The process to *Implement Action Plans* is assigned to a range of institutions, and commences with a process of review and verification of existing data sources. This contributes towards a distributed collection of selected and verified *Sectoral Information*. The selection should reflect the information needs of the CBD as defined by the selection of indicators and targets.

3.3.2 Collect Information and Fill Gaps
Depending on the extent of the information gaps identified in the *Catalog of Data Holdings*, this process may be a dominant or minor component of the overall implementation process. The additional data also contributes to building up the national collection of up-to-date sectoral information needed for estimating key indicators. The *Sectoral Information* datastores represent the main biodiversity information resource of the country. These

datastores may be extensive and held under the custodianship of a number of separate institutions. They may occur in several forms including quantitative, textual, and spatial.

3.3.3 Monitor Change
The objective of implementing action plans is to achieve positive change. Thus new information should be collected regularly, as defined in *Actions*, to keep *Sectoral Information* datastores up to date. This may involve long-term site monitoring programmes to record plant and wildlife populations, or regular habitat monitoring schemes using aerial photography or remote sensing.

3.3.4 Enact Legislation
One of the primary tools for implementing action plans is the enactment or amendment of legislation, regulations and policies (eg to create protected areas, to encourage or restrict certain human activities, to limit industrial wastes). This results in a datastore of laws, regulations and policies which is largely textual in nature, except for quantitative tables or, for example, standards reflecting regulatory limits.

3.3.5 Perform Other Actions
Although legislation and regulation are important actions, a range of other activities are desirable including:

- institutional strengthening
- human resource development and training
- monitoring and enforcement of regulation
- biodiversity research
- operational actions to reduce threats.

Many national institutions may be involved in this process, which may be broken down into a large number of sub-processes depending on specific national strategies and priorities. Such sub-division is beyond the scope of this document.

3.4 Evaluate Results
The level 2 data flow diagram for this process is illustrated in Figure 3.4.

3.4.1 Estimate Indicators
Indicators are estimated (or calculated where possible) on the basis of data held in the *Sectoral Information* datastore. This results in a further datastore of key numeric *Indicator Values*.

3.4.2 Assess Current Status vs Targets
In this process, the indicators and other results are compared to their original targets. This may involve simple numeric comparison, but more commonly, analytical assessments of progress towards targets. In addition, the effectiveness of legislation and regulation should also be assessed. The result of the analysis is a datastore of *Assessments* containing both quantitative and textual (explanatory) material.
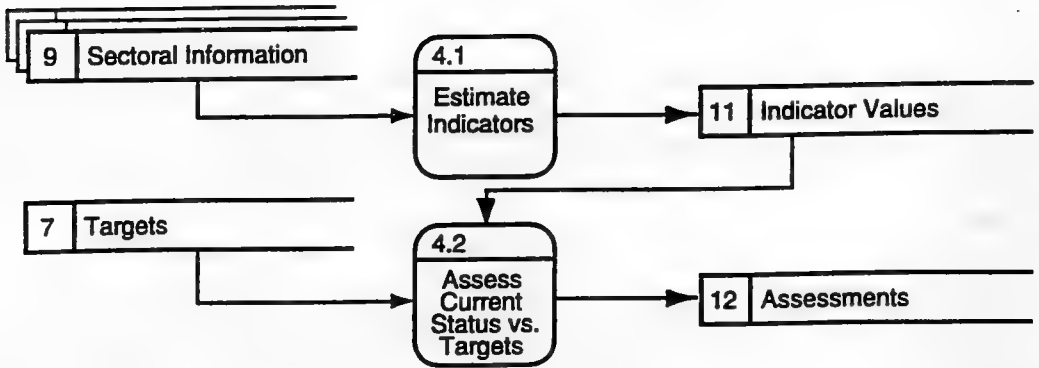
Figure 3.4: Evaluate Results (DFD - Level 2)

3.5 Report to CBD

The reporting requirements of the CBD (see Article 26) have not yet been fully defined. For this reason a detailed breakdown of the *Report to CBD* process cannot be given. However, a suggested outline of the process is shown in Figure 3.5.
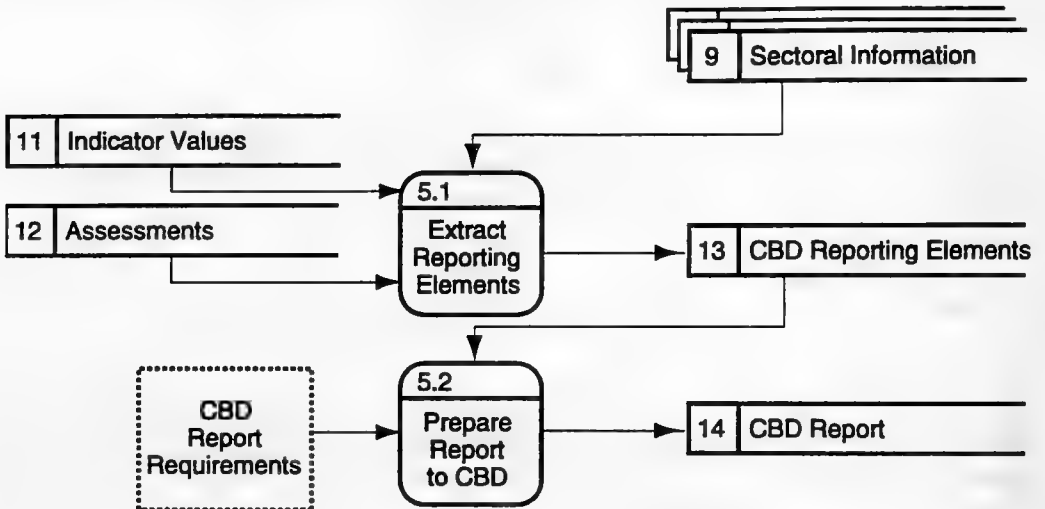

Figure 3.5: Report to CBD (DFD - Level 2)

4 DATASTORES

4.1 Overview of Datastores

The analysis of the CBD process depicted in Sections 2 and 3 indicate the presence of fourteen "datastores". As previously indicated, these datastores are conceptual, representing logical groupings of information required by or produced by the processes; no assumption is made on how and where the data may be kept. Datastores do not equate to physical datasets, data holdings or institutions; the information required for a datastore may reside in a number of institutions and derive from a range of disciplines. Note especially that datastore 9, *Sectoral Information*, represents the major national repository of scientific data relevant to biodiversity, and for this reason is depicted as multiple datastores. This section examines datastores from the perspective of data structure and content. The fourteen datastores **in total** represent the "Biodiversity Databases" identified in the level 1 data flow diagram (Figure 2.1). Each is briefly outlined, following which selected datastores are expanded in more detail in Sections 4.2 to 4.5. This initial pass at a data flow model focuses on core scientific biodiversity data. Subsequent documents are planned to incorporate other relevant domains.

1 Institutions

This datastore keeps the information about the institutional strengths and biodiversity information analysis and management capacity of the country. The custodian of this information would normally be a lead agency in the implementation of the CBD, and commonly would be implemented as a metadatabase. Wide distribution or ease of access by all other agencies is important. The process of compiling this datastore is the subject of Document 2 of this series. An important function of this datastore is to connect to the reservoir of biodiversity technology and the associated enabling technology (survey and monitoring techniques such as remote sensing), and to sources of expertise for institutional strengthening. Datastores on technology are currently beyond the scope of this model, but could be linked in to this key datastore, the structure of which is elaborated in Section 4.3.

2 Preliminary Catalog of Data Holdings
3 Catalog of Data Holdings

These two datastores could be implemented as a single evolving metadatabase. This would identify who has what data of relevance to the CBD, a key element in finding the required information for analysis as well as identifying information gaps. The structures of these datastores are elaborated in Section 4.4.

4 Human Activity and Impacts

This datastore is multi-sectoral and covers the driving forces which influence, positively and negatively, the conservation and sustainable use of biodiversity. It should include databanks on industrial and agricultural activities and outputs, population and social factors, land tenure, landuse change, etc. This encompasses much of what is often referred to as "environmental statistics". The scope of this document does not allow for the elaboration of the structure of this large and complex datastore. However, suitable data structures for environmental statistics are well established, with standard frameworks available from organisations such as the UN Statistical Office and the OECD (see *Resource Inventory*, Document 4, Section 5.9).

---

5 *Economic Values*

An assessment of the economic value of biological resources is extremely important in fostering their sustainable use and ensuring equitable sharing of benefits. Some models for organising this information have been proposed, but none are universally accepted. Implementation of such a datastore is likely to depend on national accounting and statistical systems, and will vary between countries as a result. Some guidance on assessing economic values and structuring the information is provided in the UNEP *Guidelines for Country Studies* (UNEP, 1993).

6 *Objectives*

This datastore contains national biodiversity objectives at the strategic level. These would normally be framed in general terms, eg "to sustainably harvest forest products while maintaining biodiversity". This datastore would normally be implemented in narrative form (electronic or manual) with a simple structure, such as a sectoral sub-division on the lines of the governmental program delivery structure.

7 *Targets*

This datastore identifies measurable results sought in particular time frames. Although targets might be framed in narrative terms, ideally they would be connected to quantitative indicators. Implementation should therefore be integrated with the *Indicator Values* datastore, which is elaborated in Section 4.2.

8 *Actions*

This datastore comprises the information on proposed and ongoing actions designed to achieve specific targets. This includes information on field projects, biodiversity research activities, biotechnology acquisition plans and projects, and planned and implemented policies and programmes (including those aimed at equitable sharing of benefits, sustainable use and conservation). This is not a scientific datastore, but rather an information communication and referral tool. Implementation in the form of documented "Action Plans" and bibliographic referral is most appropriate. The structure might logically parallel that of datastores 6 and 7. Guidance for organising the information of datastores 6 through 8 can be found in *National Biodiversity Strategies* (UNEP/WRI, 1994).

9 *Sectoral Databases*

As outlined in Section 1.2, the range of information which relevant to the CBD is vast. These sectoral databases refer to the observational scientific data, traditionally gathered, collected and managed in sectors such as marine science, soil science, agriculture, wildlife, botany, zoology, forestry, genetic resources, and so on. The way in which these sectors are divided varies from country to country, depending on the scientific and administrative structure. Two main classes occur within this group:

- **Core Biodiversity Data**
  Those data which relate directly to plants, animals, their habitats and ecosystems, and related genetic resources.

- **Natural Resource Base Data**
  Those data which define the resource base for biodiversity, including, soil, geology, land capability and use, climate, physiography, hydrology, and aquatic resources.

Potential data structures for core biodiversity data are elaborated in Section 4.5. The management of natural resource base data is relatively mature (compared to biological/ecological information). Thus conventions, classification systems, and standard approaches have been defined in many areas such as soil science, geology, climate, and oceanography. It is beyond the scope of this document to provide further detail of these data management conventions. However, reference is made to the work of the International Council of Scientific Unions (ICSU) CODATA program, to the practices and standards of the various sectoral scientific unions (such as, the International Society for Soil Science), and to Document 4, Sections 4 and 5.

10 Laws/Regulations
This datastore maintains information on the laws, regulations, policies, etc, which govern the use and conservation of biological diversity, including related areas which effect commercialisation, economics and benefit sharing. The structure of such a datastore is best implemented as a simple index or metadatabase providing assistance in locating relevant documents. This would be similar to any bibliographic or document management system, such as the one employed by the IUCN Environmental Law Centre.

11 Indicator Values
Indicators improve communication by quantifying and simplifying information. They can provide policy and decision makers with essential information on the status and trends in biodiversity conservation, and can help evaluate effectiveness of conservation efforts in relation to explicit management objectives or targets. They can be applied at scales ranging from community level (to guide resource managers) up to national and international levels, and can provide a framework for the collection and reporting of information at all these scales. There is a continuous need for comparison of indicators, and where possible a common approach to their selection, measurement and reporting should be introduced. This also applies to the setting of baselines and targets. The *Indicator Values* datastore is closely tied to that of *Targets*, and is elaborated in Section 4.2.

12 Assessments
Assessments are the results of analytic comparison of existing conditions (derived from monitoring) with identified targets. These normally take the form of narrative reports with quantitative tables of calculated indicator values and other summary data. This datastore could be implemented as a document management system similar to that advocated for datastore 5, 6, 7 and 11. Many countries may choose to integrate the *Targets* and *Assessments* datastores to implement the desired feedback loop of the CBD process. Quantitative tables included in the assessments could be integrated with national "State-of-the-Environment" reporting systems where these exist, using the data structures recommended by UNEP or Organisation for Economic Cooperation and Development (OECD).

13 CBD Reporting Elements
14 CBD Report

The CBD reporting elements consist largely of information extracted from other datastores, especially from the *Sectoral Information*, *Actions*, and quantitative component of the *Assessments* datastore. As the reporting requirements for the CBD are not yet fully defined, it is not appropriate to explore a data structure for these datastores. However, the quantitative component of the *Reporting Elements* datastore is likely to be maintained as a set of relational database tables linked to narrative assessments in the form of text-based files (for guidance on this approach see Document 3, Section 3.6).

4.2   Indicator Values

Indicators are determined by the specific issue under consideration, the target users of the indicator, its spatial and temporal scope, data availability and the framework available for analysis. They should be relevant to policy, should be well founded in technical and scientific terms, and should be measurable. They may have a range of components and may draw on data contained in several datastores such as economic values, human impacts and sectoral databases (eg protected areas, habitats, physical features). Indicators are "information" derived by analysing primary data, even though they may be presented in tabular, map, or graphical form. Their value derives from their ability to place data in the context of agreed baselines and targets.

Many groups have developed indicators for environmental, social and economic monitoring, and new indicators are continually being developed for application at sectoral, national and even global scales. The World Bank, Organisation for Economic Cooperation and Development (OECD), and the World Resources Institute (WRI) are but a few of these. WCMC is currently testing a number of indicators of forest condition in a series of case study sites in tropical regions.
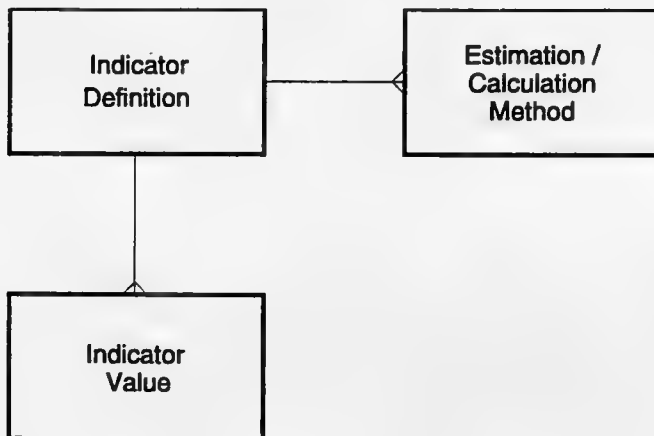

Figure 4.1: Indicator Values Data Model

However, there continue to be problems in the development of consistent measurable biodiversity indicators. These mainly stem from a lack of appropriate primary data, but are also subject to debate over definitions, inadequate comparability of baselines and goals. As a result there is no universally accepted data structure typifying a datastore of *Indicator Values*.

Figure 4.1 suggests a simplified structure for such a datastore, with three entities: the **indicator value** (specific instances of the measurement or estimation of an indicator at a particular time and place); the **indicator definition** (which may be both descriptive and quantitative and carry attributes such as the desired level of the indicator, or its critical threshold); and **estimation/calculation method** (there may be more than one acceptable estimation method for a given indicator).

4.3 Institutions

An **Institution** is defined as a recognisable organisation that maintains or uses information of relevance to the CBD. The resulting metadatabase contains information on the strengths, capacities and data holdings of each institution in the country and, if relevant, region. An example of an **Institution** metadata entry is given below:

Name:	World Conservation Monitoring Centre
Acronym:	WCMC
Type:	Non-governmental
Theme:	Information Services
Keywords:	biodiversity; conservation; information
Postal_Address:	219 Huntingdon Road
Postal_Code:	CB3 0DL
City:	Cambridge
Country:	United Kingdom
Contact_Person:	Jo Taylor
Contact_Status:	Information Officer
Telephone:	44-1223-277314
Fax:	44-1223-277136
Email:	Internet: info@wcmc.org.uk
Update_Date:	1994-09-01
Mission:	To provide research, information and technical services so that decisions affecting the conservation and sustainable use of biological resources may be based on the best available scientific information.

A few sample field definitions are given below to give a flavour of the institutional metadata (the full Metadata Data Dictionary is defined in Document 2, Annex 6).

Name:

Definition	Official name of the institution.
Format	Maximum 50 characters.
Status	Mandatory.
Example	World Conservation Monitoring Centre.

Type:

Definition	The organisation type, selected from one of the following: governmental; non-governmental; commercial; academic; inter-governmental; United Nations.
Format	Maximum 20 characters.
Status	Mandatory.
Example	Non-governmental.

Theme:

Definition	The primary function of the organisation, selected from one of the following: research; consultancy; information services; campaigning. The selection of a primary function keyword is not intended to wholly define the scope of the organisation. Detailed description of the function of the organisation can be expanded on in the "Mission" section.
Format	Maximum 30 characters.
Status	Mandatory.
Example	Information Services.

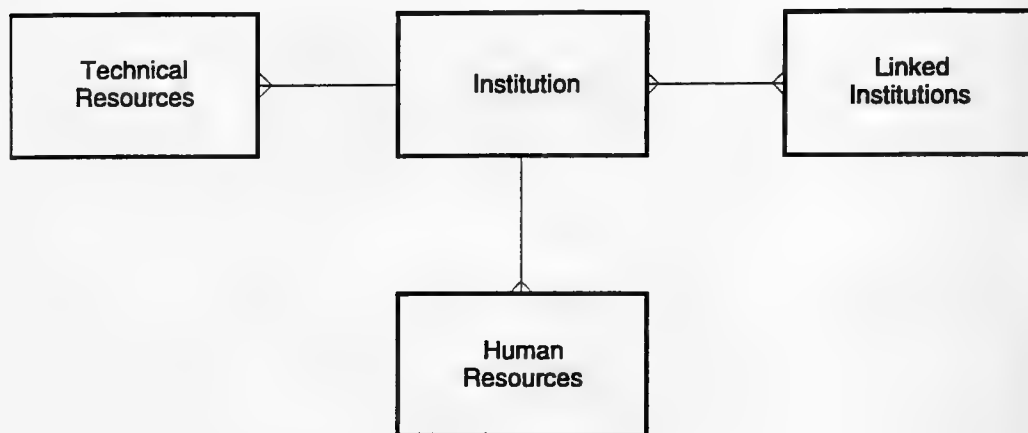A data model for this metadatabase is shown in Figure 4.2.

Figure 4.2: Institutions Data Model

4.4   Catalog of Data Holdings

The **Dataset** is defined as a collection of data and accompanying documentation maintained at an **Institution**. A collection of data refers to one or a series of Data Members which relate to a specific *theme* or *geographic region*. Sample definitions of **Dataset** metadata items are:

Name:

Definition	The name given to the dataset or activity being described. The title should be descriptive enough to allow the reader to make a reasonable decision as to whether the data may be of interest.

Format	Maximum 50 characters.
Status	Mandatory.
Example	African Protected Areas GIS.

Theme:

Definition	This is the theme or parameter being measured by the dataset. The keyword entered is the most general, and should, if possible, be taken from the standard terminology lists.
Format	Use INFOTERRA terminology list.
	Maximum 31 characters.
Status	Mandatory.
Example	Terrestrial ecosystems.

A data model for this metadatabase is shown in Figure 4.3.
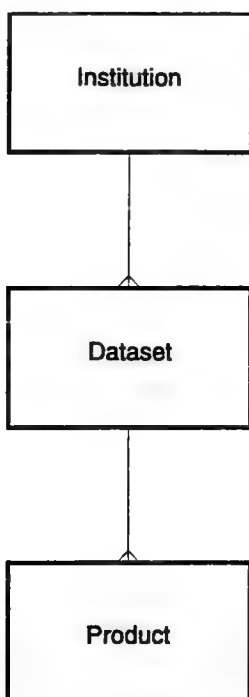

Figure 4.3: Catalog of Data Holdings Data Model

4.5   Sectoral Information - Core Biodiversity Data

4.5.1 Overview

Annex I of the CBD contains the following indicative list of categories for identification and monitoring:

"1.   Ecosystems and habitats: containing high diversity, large numbers of endemic or threatened species, or wilderness; required by migratory

species; of social, economic, cultural or scientific importance; or, which
are representative, unique or associated with key evolutionary or other
biological processes.

2.   *Species and communities which are: threatened; wild relatives of*
      domesticated or cultivated species; of medicinal, agricultural or other
      economic value; or social, scientific or cultural importance; or importance
      for research into the conservation and sustainable use of biological
      diversity, such as indicator species.

3.   *Described genomes and genes of social, scientific or economic*
      importance."

Based on this list (particularly items 1 and 2) and given the ways in which biodiversity data
are currently organised by the key agencies, the core data may be considered as relating to
four primary entities: **habitats, protected areas, species** and **threats**. These four entities
represent the first level entity-relationship (E-R) model for biodiversity data in the datastore
Sectoral Information. The relationships between these core biodiversity entities are shown
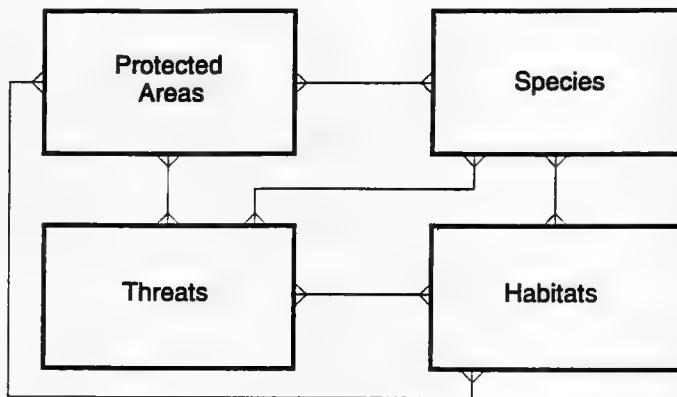in Figure 4.4.

Figure 4.4: Core Biodiversity Data Entities

The figure shows many-to-many relationships between all entities, ie:

- a **species** may be subject to many **threats**
- a **threat** may apply to many **species**
- a **species** may be in several **protected areas**
- a **protected area** may harbour many **species**.

The entities shown are at a high level (Level 1) and each can be further broken down (Level
2). These are elaborated in turn in E-R diagrams in subsequent sections. Appropriate models
for genetic resources (item 3 above) have not been developed in this document and are
planned for future research. Some work has been done on establishing effective ways of
managing agricultural germplasm information at the International Plant Genetic Resources
Institute (IPGRI) in Rome, and there are close links to the way in which species information

in general are organised (see Section 4.5.3).

4.5.2 *Habitats*

Figure 4.5 shows a structure which could be used for handling data relating to habitats. There are three entities:

- **area** is a spatial entity defining the geographic location of the habitat

- **habitats** has the basic attributes of the polygon, eg identifier, type, etc.

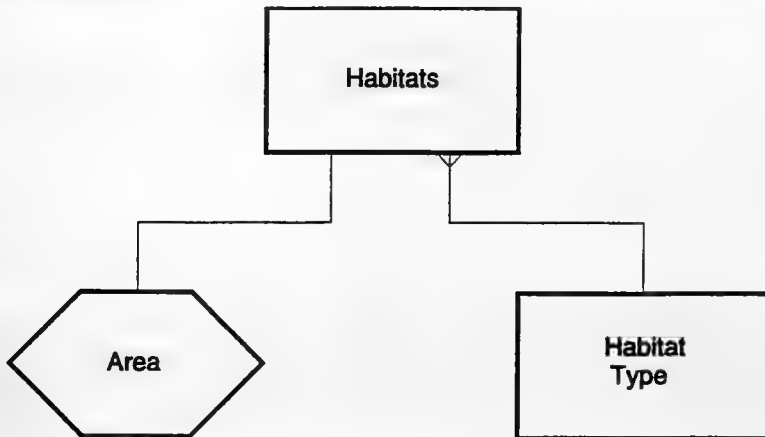- **habitat type** gives more detail of the meaning of "type", eg type, description, etc.

Figure 4.5: Habitats Data Model

There is a one-to-one relationship between **area** and **habitats**, and a one-to-many relationship between **habitat type** and **habitats**.

This structure is simple and can be implemented using elementary geographic information systems and data management tools. These allow production of maps showing the current geographical distribution of the various habitat types, with tables giving the area and, for example, percentage coverage of the total area of the country by a specific habitat type. This type of output is useful in summarising data for decision makers for planning purposes.

The structure is easily extended to cope with habitat monitoring, by maintaining a sequence of date-stamped editions describing the situation with respect to habitat at different points in time. By comparing these, maps and tables showing change (either decrease or increase) in the various habitat classes can be produced.

The principal problem in dealing with habitat data is not one of complexity in data structure at this level, but the absence of an internationally accepted habitat or ecosystem classification at an appropriate scale for national biodiversity planning. Again, the varying requirements of different countries means that a widely applicable classification is difficult to conceive. However the data management structure suggested here is independent of this problem.

This basic structure is used in the Tropical Forests Database developed by WCMC. The source data have been derived from a variety of sources including satellite imagery, existing databases, maps, survey data, and so on, and harmonised into standardised broad forest categories. For example, the four major categories of forest are *lowland rain forest*, *montane rain forest*, *inland swamp forest* and *mangrove*. The forest polygons are held in an Arc/INFO coverage with linked database tables, echoing the structure illustrated in Figure 4.5.

4.5.3 Protected Areas
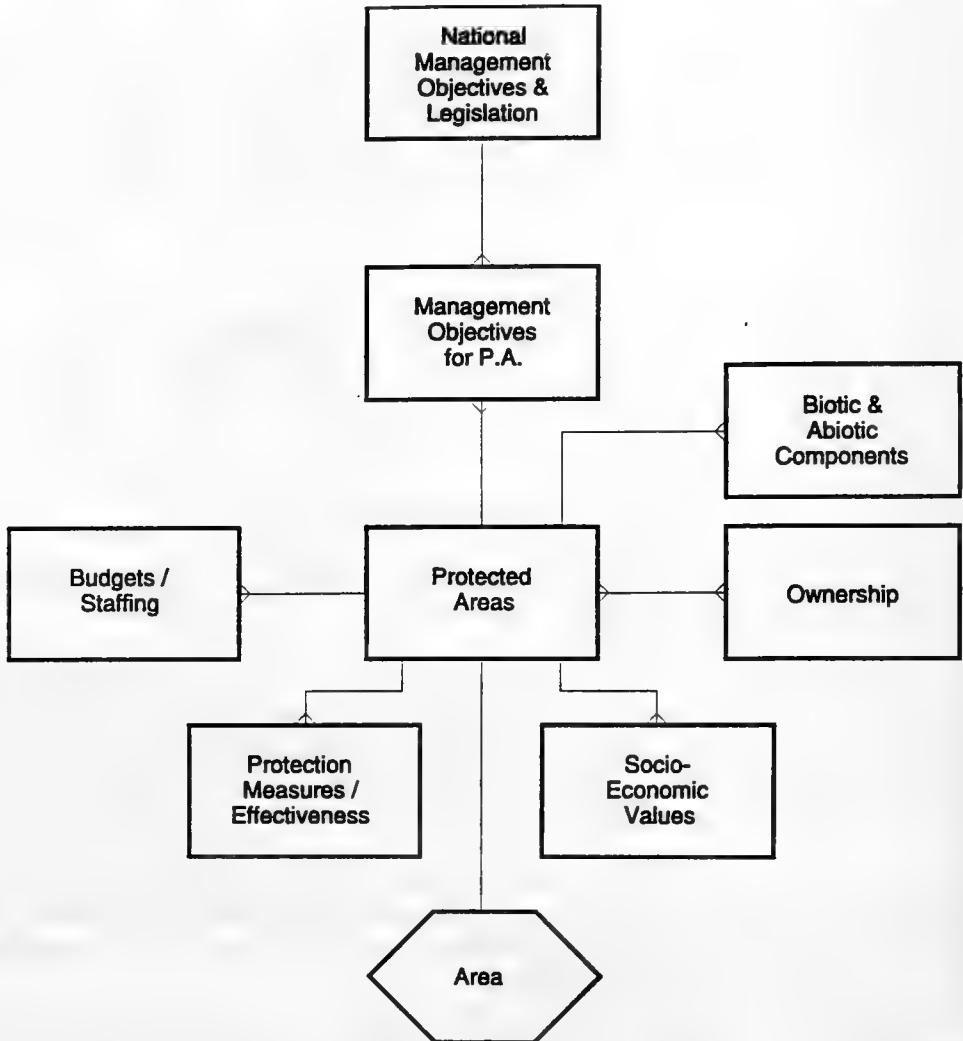Figure 4.6 shows a structure which could be used for handling data relating to protected areas.

Figure 4.6: Protected Areas Data Model

The entities and relationships are:

- **protected area** is the basic entity containing primary attributes such as name, year established, size, designation, description, etc.

- **area** is a spatial entity defining the geographic location of the protected area

- **socio-economic values** (such as tourism) may be associated with a protected area

- the land (of the **protected area**) may be owned (**ownership** entity) by one or more agencies (or individuals)

- **management objectives** will be set for the protected area (and these may relate in turn to **national management objectives**)

- the **protected area** will be assigned **budget/staffing** for its operation

- **protection measures** are established for a **protected area** and the **effectiveness** recorded.

The spatial element in this remains simple, but there are more entities than in the case of habitats and there is variation in the nature of the entities. For example, items in **budget** and **ownership** are fairly apparent, but the attributes to be included in **economic values** may be less so. The form of the entities could imply that more capability is required in the data management tools.

With a protected areas database of this form, maps and reports could be produced:

- showing the various areas under protection
- summarising the current management objectives
- giving budgetary roll-ups of expenditures on protected areas
- totalling costs and benefits of tourism (an economic value)
- highlighting interactions with surrounding land use.

There is a well established database of Protected Areas in use at WCMC. Although it includes the elements mentioned above (among others), the purpose is to provide a source of information on the world's protected areas, rather than to provide a mechanism to assist in managing protected areas which is left to experts in the countries concerned. This inevitably leads to a difference in perspective. However, some outputs similar to those above can be produced from the WCMC system, which is implemented using the FoxPro relational database management system (RBDMS) and Atlas-GIS mapping package (see Document 4, Section 3.2).

4.5.4 Species
Figure 4.7 shows a structure which could be used for handling data relating to plant and animal species.
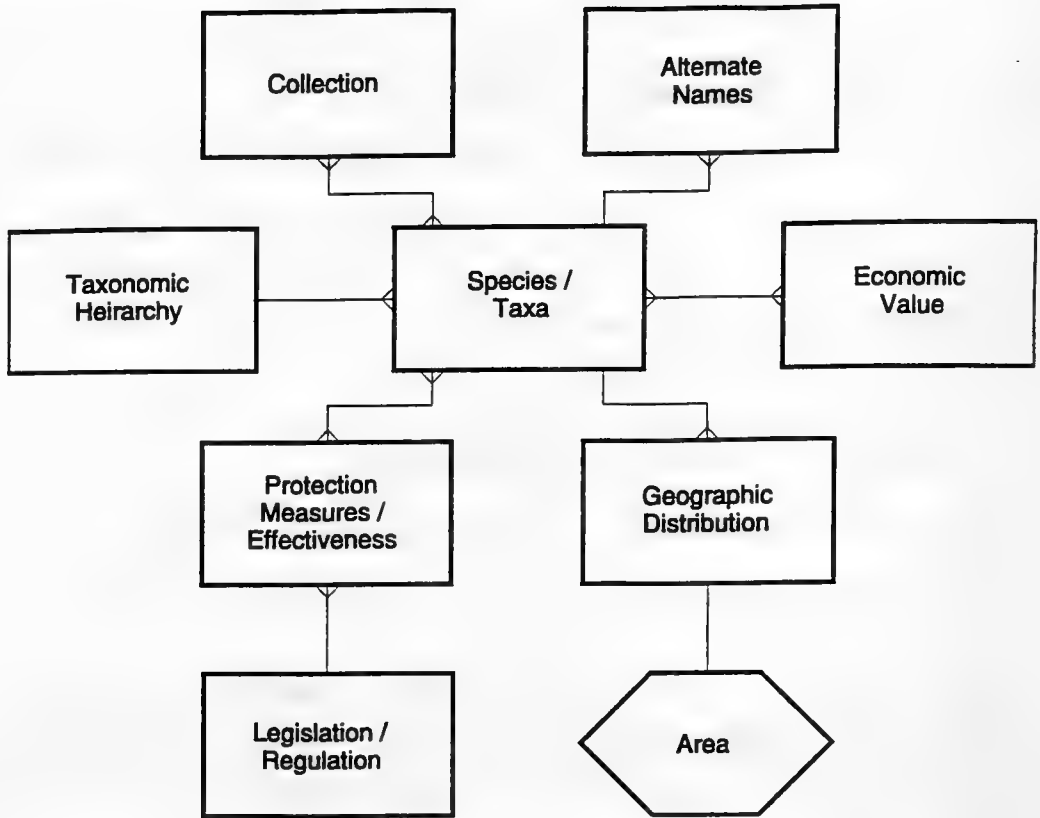
Figure 4.7: Species Data Model

The following points should be noted:

- the central entity **species/taxa** contains the basic attributes of the plant or animal, eg name, and this may be at the species or taxa level

- any **species/taxa** may have **alternate names**, both common and scientific

- more than one **collection** may relate to any **species/taxa**

- the **taxonomic hierarchy**, shown as one entity for the sake of simplifying the figure, is composed of multiple entities

- any **species/taxa** may have **protection measures** applied to it

- the **protection measures** in turn may relate to **legislation/regulation**

- all **species/taxa** are located geographically (**geographic distribution**)

- the **geographic distribution** may be located as one or more specific areas (or points)

- any **species/taxa** may have an **economic value** attached to it (which may be expressed as several entities).

Again the spatial element remains simple and the complexity is in the non-spatial attributes (especially as several of the entities are likely to expand into multiple entities) and correspondingly more complex data management tools may be required to implement such a database. In fact the form of the model is likely to be influenced by the facilities available in the computer system used.

The type of analyses which could be generated include:

- lists of species, grouped for example by family, with maps showing their distribution

- lists of endemic species and associated distribution maps

- lists of economically significant species, identifying the type of value

- cross-referencing of legislation with the species covered and the nature of their protection

- summaries of total species, number endemic to the country, species populations

Again, comparison of the results of similar analyses conducted over a period of time, permits the monitoring of species distribution and numbers.

This is a very simplified view, and additional linkages to threats, protected areas, trade in species, etc will be needed. Figure 4.8 shows, for purposes of example, the main files in the database used to manage plant information at WCMC and other establishments.
This structure echoes the data model given above, for example:

- the *names* table is the central species/taxa entity

- the *genera*, *families*, *orders*, *subclasses* tables give the taxonomic hierarchy

- the *distributions* table links to *WCMC areas* (Biodiversity Reporting Units) providing an index to geographic location

- the *distributions* table also contains a conservation status item which may link to *laws* (c.f. protection measures and legislation).

Note that the E-R diagram in Figure 4.8 showing the general relationships between the principal data files in the database is diagrammatic, but not strictly correct.
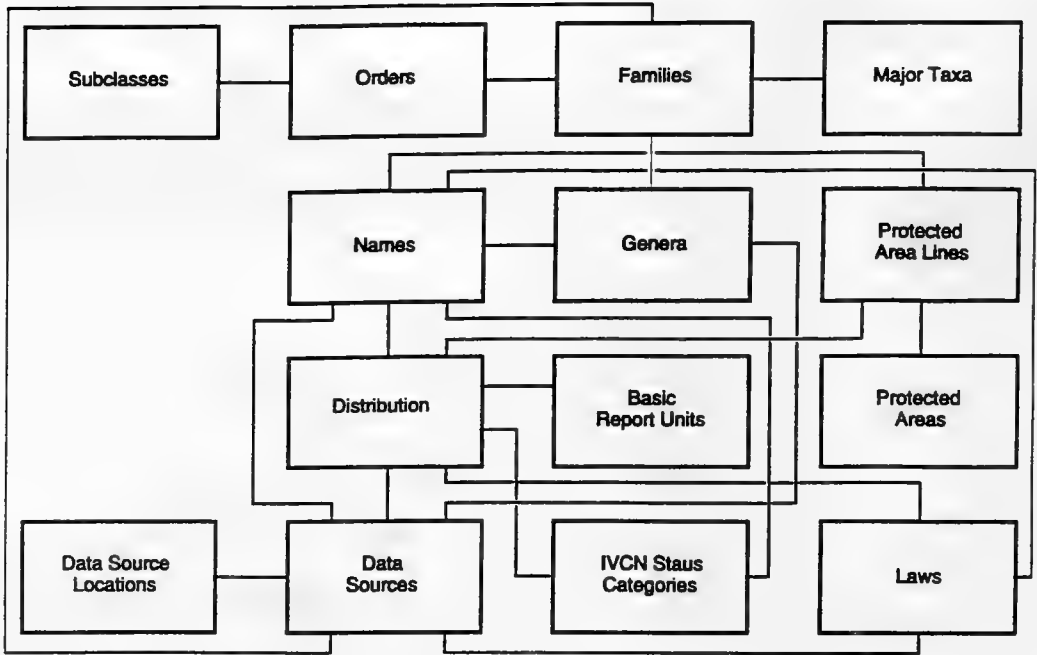
Figure 4.8: Main Elements of WCMC Plant Database (*"BG-BASE"*)

Note also that the conceptual E-R diagram of Figure 4.7 contains no parallel to the "data sources" and "data sources location" table. These are linked to many other tables and provide a mechanism for documenting the source of the various items of data. This is valuable information when dealing with biodiversity data of all kinds, and arises not so much from the structure of the information itself, but from the requirements for managing it.

This particular implementation uses a system called *BG-BASE*, which has been developed using the Advanced Revelation, a relational type of database management system (RBDMS) which allows for variable field lengths and multi-value fields.

4.5.5 Threats
The data which are needed to describe and analyse threats, and the structures needed to effectively manage them, depend heavily on the nature of the threat. For example if the threat to a species originates from trade, then details of the traffic and trade in related commodities would be relevant; if the threat is due to loss of habitat, then rates of land use change and related spatial information must be recorded. Different data structures are required in each case. Threats deriving from widely distributed phenomena such as climate change or long-range transport of pollutants, present different challenges again for data organisation.

While the Country Studies Guidelines (UNEP, 1993) outline three major classes of threat: *External Socio-Economic Factors*, *Direct Threats: Local Impact*, and *Direct Threats: Global Impact* - in designing a national program of information management on threats, emphasis should be placed mainly on the proximate (or "direct") threats of local origin. Socio-economic factors should be considered as causal factors, rather than as threats *per se*.

This distinction is not always clear. There is considerable interrelation between causal factors, threats to species, threats to habitats, human activities, mitigation measures, etc. "Threats" to one species may result from measures aimed at conserving another. It is a complex situation without as yet a great deal of standardisation of concepts and approaches. Listed in UNEP (1993) are seven major categories of human-induced threats with very many sub-classes, and a number of other ways of categorising threats are available.

From the perspective of threats to species, a primary breakdown would distinguish between threats to the habitat of the species (such as loss, fragmentation and degradation of habitat quality), and threats to the species itself (such as harvesting, hunting, introduction of competitive species). The E-R diagram which follows (Figure 4.9) applies to the organisation of information on threats which are internal to the country. Different structures will be required to deal with external threats such as climate change.
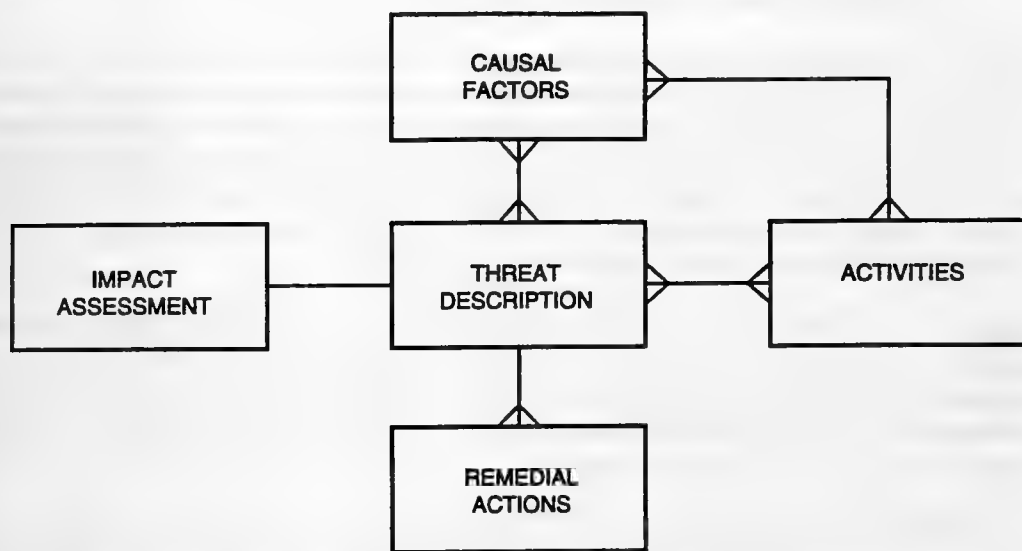

Figure 4.9: Threats Data Model

Referring to Figure 4.9:

- The central entity **threat description** contains data such as threat category (preferably following the standard IUCN classification), intensity and duration, a narrative description of the threat, and pointers to relevant references.

- the **activities** entity records the quantitative information related to the threat, for example how much, how many, where, etc. The exact structure of this will be highly dependent on the nature of the threat. For instance if the threat was "road building", activities might include the length, nature and position of roads, associated support facilities, construction timetables, and so on. If "hunting" was the threat, then records of annual take and the nature and number of hunting parties might be relevant. Many activities are likely to be associated with each threat and *vice versa.*

- The **causal factors** entity identifies and describes the primary driving forces which generate threats. For example, in the case of "road building", causal factors might be mining and tourism. Socio-economic and other human factors (see suggested list in UNEP, 1993) would also be listed here. Many causal factors may relate to each threat and *vice versa.* Similarly, causal factors may generate many specific activities which need monitoring.

- The **remedial actions** entity includes data on the feasibility of actions to reduce the threat, as well as the costs and benefits involved.

- The **impact assessment** entity contains estimates of the likely effects of the threats, both ecological and economic, and suggests their potential for reversal.

This data model breaks down the **Threats** entity of the first level model of Figure 4.4, but is still aimed at a general and generic level, and is thus provided as an example. Expanding the model to include threats to habitats will require at the very least the addition of a spatial entity to reflect the geographic extent of the threat. Other additions which should be considered are linkages to information on institutions and their capabilities and roles in remedial actions, traditional uses of the threatened habitats, economic benefits of resource utilisation, international conventions and treaties, and so on.

Expanding or modifying the model to deal effectively with non-proximate threats, and more general causal factors (such as marine pollution driven by industrial development, and ultimately by population pressure) requires the introduction of additional entities such as "ultimate driving forces", and "regional (or collective) threats" which would hold information which is not specific to a species or habitat.

The consideration and classification of threats is at an early stage (see Document 4, Section 5.9), so no example can be given of an operational database based on this model. At the present time, threats are normally described in narrative form in species or protected areas information systems.

4.5.6 *Integrating Core Biodiversity Data*
Existing (computerised) databases tend not to attempt to hold all biodiversity information in a single structure. The focus in any one implementation tends to concern a particular entity (or small group of entities), with investment in **detailed** data about that entity. This may be due to the sectoral mandate of the agency implementing the database, lack of data, or because the user requirements are successfully achieved with that limited information. This is not to say that relationships to other entities are necessarily ignored. For instance, both the

examples of species and protected areas databases given above contain data w threats. Arguably this is minimal within the databases, but users could establish links to other (perhaps non-computerised) databases containing further details if desired.

In the context of the CBD, some countries may wish to integrate all their biodiversity information into a single information system at one site, such as a national environmental information centre. For the reasons mentioned above, others may prefer to leave the responsibility of data custodianship to several agencies, and implement an effective coordinating mechanism. The latter could be the foundation of a phased approach leading to integration in the longer term. Regardless of the mechanism, the work to produce standards for classification schemes, agreed taxonomies, data transfer mechanisms, and high level dataflow models, is needed even where custodianship is maintained on a sectoral basis, in order to have a *conceptually* integrated biodiversity information management system for the country. This integrated view will then permit the development of sound national strategies and actions in response to the CBD.

5 REFERENCES

Ashworth, C. and Goodland, M. 1990. *SSADM: A Practical Approach*. McGraw Hill.

Chen, P. 1976. The Entity-Relationship Model - Towards a Unified View of Data. *ACM Trans. on Database Systems*. 1:9-36.

Yourdon, E. 1979. *Structured Design: Fundamentals of a Discipline of Computer Program and Systems Design*. Prentice Hall.

UNEP 1993. *Guidelines for Country Studies on Biological Diversity*. United Nations Environment Programme, Nairobi, Kenya.

UNEP/WRI 1994. *National Biodiversity Strategies - Guidelines for Biodiversity Planning and Profiles from Early Country Experience*. World Resources Institute, Washington DC, in prep.

WCMC 1994. *The Biodiversity Clearing House - Concept and Challenges*. WCMC Biodiversity Series No 2, World Conservation Press, Cambridge, UK, pp.34.

WRI/IUCN 1993. *Biodiversity Indicators for Policy Makers*. World Resources Institute, Washington, DC, pp.42.

ANNEX 1: ANALYSIS OF THE INFORMATION NEEDS OF THE CBD

Article	Information Requirements	Type[*]
1. Objectives	None	
2. Use of Terms	None	
3. Principle	How "actions within the jurisdiction" are affecting the environment of other jurisdictions.	Informative
	Global and regional state-of-the-environment information.	Scientific
	Interrelationships and effects of outputs (pollution, wastes etc).	Scientific
	How trading and other economic activities affect the environment of other nations.	Economic
4. Jurisdictional Scope	None	
5. Cooperation	National institutional strengths and capabilities.	Metadata
	Needs of Contracting Parties.	Informative
	Strengths and capabilities of "competent international organisations".	Metadata
6. General Measures for Conservation and Sustainable Use	Current sectoral and cross-sectoral plans and strategies which may effect conservation and sustainable use of biodiversity.	Informative and metadata

7. Identification and Monitoring	[As identified in Annex I to the CBD]:	
	1. Ecosystems and habitats: containing high diversity, large numbers of endemic or threatened species, or wilderness; required by migratory species; of social, economic, cultural or scientific importance; or, which are representative, unique or associated with key evolutionary or other biological processes.	Scientific
	2. Species and communities which are: threatened; wild relatives of domesticated or cultivated species; of medicinal, agricultural or other economic value; or social, scientific or cultural importance for research into the conservation and sustainable use of biological diversity, such as indicator species; and	Economic, social Scientific
	3. Described genomes and genes of social, scientific or economic importance.	
	A broad program of systematic data collection on species, protected areas, critical habitats, and ecosystems (these are referred to as the "core biodiversity data" and elaborated considerably in the body of this document).	
	Human activities, including industrial activities, agricultural practices, land use etc.	
	Monitoring data on the state of the environment, measured according to standard procedures in continuous time series.	

8. In-Situ Conservation	8a to h As in Article 7, but with more specific reference to protected areas and to the status of ecosystems and species populations.	Scientific
	8g The effects of the introduction of "living modified organisms" in other jurisdictions, and monitoring data on local effects.	Informative and scientific
		Scientific
	8h Information on alien species (presence, sources).	Informative
		Informative
	Eradication and control measures for alien species.	Informative
	8i Present uses of biodiversity.	Economic
	8j Innovations and practices of indigenous and local communities.	Legal
	Accrued benefits (especially economic) of use of biodiversity and relative contributions of local communities (to enable fair sharing of benefits).	Legal
	8k Existing legislation and regulation on protection of endangered species.	Economic
	Effects of legislation and regulation in other jurisdictions.	
	8l,m Financial requirements of conservation measures.	

9. Ex-Situ Conservation	Capabilities and facilities of institutions for research and ex-situ conservation.	Metadata
	Research results on effective methods of re-introduction of species.	Informative
10. Sustainable Use of Components of Biological Diversity	**10a,b** As per Article 6.	Informative
		Informative
	10c,d Customary use and traditional cultural practices and how these can be used for remedial action.	Metadata
	10e Strengths and capabilities of private sector organizations for the development of methodologies.	
11. Incentive Measures	Incentive measures found to be effective in other jurisdictions.	Informative
	Cost/effectiveness of incentive measures employed.	Economic
12. Research and Training	Training and education needs and priorities.	Informative
		Metadata
	Available sources of training and education.	Metadata
	Biodiversity research activities world-wide.	
13. Public Education and Awareness	Available materials suitable for public awareness.	Metadata
	Successful awareness tools and activities.	Informative
	Bibliographies, technology of networking and information exchange tools.	Metadata, informative

14. Impact Assessment and Minimising Adverse Impacts	Major projects which may have impact on biodiversity.	Metadata
	Impact assessment methodologies.	Informative
	Resources and population at risk in-country and in neighbouring regions.	Scientific
	Nature, availability and location of emergency response facilities.	Informative
	Emergency response contingency plans and strategies.	Informative
15. Access to Genetic Resources	**Pts 1-6** Systematic record of available genetic resources (germplasm, plant and animal genetic research results).	Scientific
	Environmental sound uses of genetic resources.	Informative
	Pt 7 Benefits, commercial and otherwise, of genetic research and resulting genetic resources.	Economic
16. Access to and Transfer of Technology	As in Article 15 but related to technology innovation rather than genetic resources.	Scientific and informative
17. Exchange of Information	Bibliographies, directories, metadatabases on research, technology, and available data (world wide).	Metadata
18. Technical and Scientific Cooperation	National institutional strengths and capabilities.	Metadata
	Technical and scientific advances and research programmes of Contracting Parties.	Metadata

19. Handling of Biotechnology and Distribution of its Benefits	[From Article 19] "*any available information the use and safety regulations required by that Contracting Party in handling such organisms, as well as any available information on the potential adverse impact of the specific organisms concerned to the Contracting Party into which those organisms are to be introduced.*"	Legal
20. Financial Resources	Financial resources available to support activities under the CBD.	Economic
	Economic and social conditions within the developing countries.	Social
	Environmental conditions within the developing countries.	Scientific
21. Financial Mechanism	As per Article 20.	Economic
22. Relationship with Other International Conventions	Terms and conditions of other relevant international conventions.	Legal
23. Conference of the Parties	None	
24. Secretariat	None	
25. Subsidiary Body on Scientific, Technical and Technological Advice	Integrated information from all other Articles.	Scientific, informative
26. Reports	Yet to be determined.	
27-42.	Procedural and administrative Articles with little information management requirement.	

°Types of Information:

Informative. Descriptive information about the issue. Usually in narrative form.

Legal. Regulations, legislation and other legal instruments.

**Scientific.** Measured or scientifically observed data, often in numeric or categoric form in databases.

Economic. Information related to costs, expenditures, and other financial information, usually in numeric form.

Social. Information on population, health, and other social measures, usually in numeric form.

ANNEX 2: LIST OF ACRONYMS & ABBREVIATIONS

BDM	Biodiversity Data Management
CBD	Convention on Biological Diversity
DFD	Data Flow Diagram
ERD	Entity Relationship Diagram
GEF	Global Environment Facility
GIS	Geographic Information Systems
ICSU	International Council of Scientific Unions
IUCN	World Conservation Union
OECD	Organisation for Economic Cooperation & Development
UNEP	United Nations Environment Programme
WCMC	World Conservation Monitoring Centre
WRI	World Resources Institute

NB See also the index of acronyms and abbreviations in the *Resource Inventory* (Document 4).
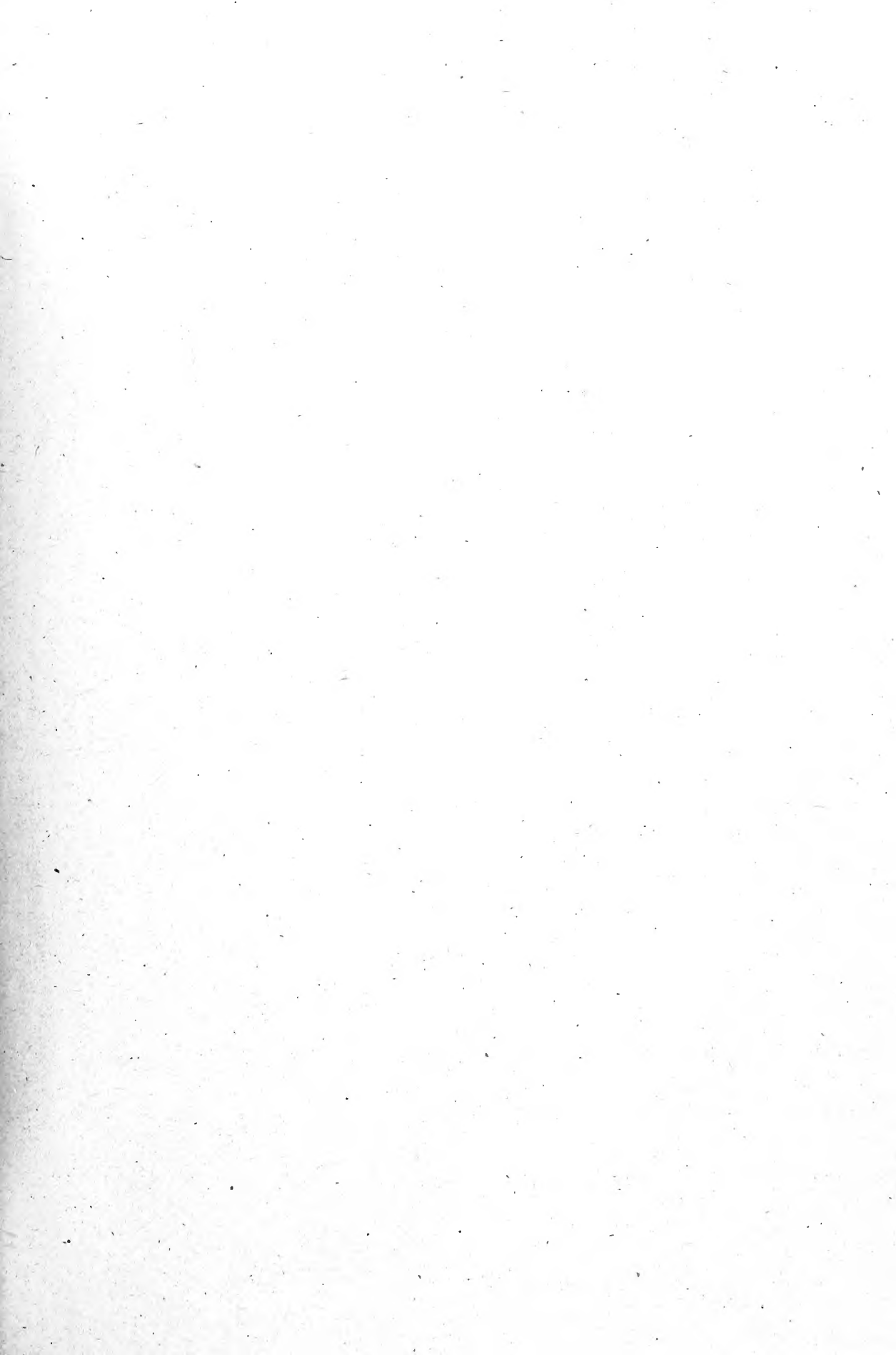
WORLD CONSERVATION MONITORING CENTRE

World Conservation Monitoring Centre
219 Huntingdon Road
Cambridge CB3 0DL
United Kingdom

Telephone +44 223 277314
Fax +44 223 277136