Biodiversity Data Management Project

FRAMEWORK FOR INFORMATION MANAGEMENT

in the Context of the
Convention on Biological Diversity

United Nations Environment Programme

in association with the
World Conservation Monitoring Centre

UNEP

WORLD CONSERVATION
MONITORING CENTRE

Second Edition

23094

UNEP - United Nations Environment Programme is a secretariat within the United Nations which has been charged with the responsibility of working with governments to promote environmentally sound forms of development, and to co-ordinate global action for development without destruction of the environment.

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.

For additional copies of this document
or further information please contact:

Task Manager
Biodiversity Data Management (BDM) Project
United Nations Environment Programme
PO Box 30552
Nairobi
KENYA

TABLE OF CONTENTS

TABLE OF FIGURES

ACKNOWLEDGEMENTS

This document is one of a series researched and compiled by the World Conservation Monitoring Centre, Cambridge UK with 80% funding from the Global Environment Facility through the United Nations Environment Programme, Project GF/0301-94-40 (GF/0301-94-06). The remaining 20% funding was provided by WCMC. The need for the development of a package of tools and materials to support national information management for the CBD was identified jointly by UNEP and WCMC.

The first edition of this document *Guidelines for Information Management* (March 1995) was prepared under the editorship of Ian Crain. This revised edition was prepared by Jake Reynolds and John Busby.

The editors wish to thank Allen Hammond for advising on many of the concepts in this document, Feargal Duff for managing and contributing to the document revision process, the Advisory Panel of the Biodiversity Data Management Project for their support and suggestions, and the staff and consultants of WCMC for their valuable contributions.

PREFACE

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 organize, by any mechanism, data derived from identification and monitoring activities".

In response to this requirement, a project was initiated by the United Nations Environment Programme and World Conservation Monitoring Centre to facilitate the building of national capacity for biodiversity data management and exchange as required by the CBD. One of the outputs of the GEF-funded *Biodiversity Data Management* (BDM) project is a set of supporting materials intended to raise the profile of biodiversity information in decision-making processes, and help countries establish information programmes in support of national biodiversity strategies and action plans. The materials, which were prepared by WCMC, comprise: *Framework for Information Management* (this document), *Guidelines for National Institutional Survey*, and the *Electronic Resource Inventory* (UNEP/WCMC 1995).

This document, covering a wide spectrum of information issues, is divided into seven chapters. **Chapter 1** reviews the role of information in decision-making, emphasising the importance of continuous improvement in biodiversity planning. **Chapter 2** considers the organisational issues surrounding multi-agency information system development, covering topics such as organisational structure, co-ordination and priority setting. **Chapter 3** discusses methodologies for building information systems within an agency or groups of agencies, emphasising the need for user needs assessment. **Chapter 4** examines key data management concepts, including the use of primary data, standards, and formal database development techniques. **Chapter 5** tackles the issue of quality management, examining institutional quality procedures, dataset documentation, operational and data security, and human resources. **Chapter 6** concentrates on information production, presenting ways of integrating, analysing, and delivering information to different audiences. Finally, **Chapter 7** comprises a list of **references** cited in the text, and a **glossary** of biodiversity and information management terms.

There is no single way to achieve improvements in the environment through the use of information. In all cases the approach has to be tailored to local conditions. The intent here is to provide a reference work of broad scope to act as a framework for agencies and individuals implementing their own priorities for information management.

1

Information for Decision Support

1.1 INTRODUCTION

The world's biological resources are rapidly being degraded due to unsustainable human activities. The changes which are occurring threaten the long-term survival of many lifeforms, including our own. Particularly serious impacts for humankind include the erosion of genetic variability, the decline in health and functioning of ecosystems, the compromise of food and water security, and the emergence of lethal micro-organisms.

The major forces behind these impacts are technological innovation, economic development and population growth (A.Hammond, pers. comm.). Not only have these occurred with increasing speed during the twentieth century, they have done so in an uneven pattern across the globe. This has led to additional stresses and divisions between nations and smaller groups as they compete for scarcer resources.

Biodiversity is now a concern to all sectors of society. Individuals, local communities, industry, sovereign states and international organisations all make decisions which affect the sustainability of biological resources. One response to the worsening situation is the *strategic* use of information. Groups around the world are recognising that organised information is empowering, and are taking the necessary steps to reorient their activities in favour of effective information management.

Useful information has certain characteristics: it is relevant to the decisions being taken; it is timely, in that it is available when and where it is needed; and it can be interpreted easily without special training or technology. Such information can be absorbed into the decision-making process and counteract the current environmental decline.

Many groups already possess information, of a cultural or scientific nature, which is of great value to others. However, the exchange of information between different levels and groups in society is frequently restricted and, in their broadest sense, information systems are intended to facilitate this exchange.

Biodiversity issues tend to be complex, involving a large number of stakeholders with widely differing perspectives and needs. Simple answers to complex questions are often incorrect, and can generate new problems themselves. Indeed, biodiversity issues are frequently obscured by a 'tyranny of small decisions', without anybody taking responsibility for reconciling different points of view. Difficulties in defining the value and benefits of biodiversity are well stated in the Technical Annex to the *Guidelines for Country Studies on Biological Diversity* (UNEP 1993):

> "Countries will find that their efforts to measure the value of biological resources and diversity are hampered by tremendous uncertainty. There is uncertainty regarding biological measures of the qualities, quantities, diversity and interactions of biological resources. There is uncertainty of the various goods and services that flow to us from these resources, or that may flow to us in the future. There is also uncertainty about the values members of our society place upon the flows of these goods and services and the values that future generations may place upon them. There is uncertainty about how human actions may impact biological resources and diversity and their associated goods and services, but we face the very real risk that the impacts of our actions may be irreversible. This is clearly the case for extinction of a species due to unsustainable use or disruption of habitat".

National information responses to this crisis tend to follow a similar path. The initial push is from community groups, professional associations, and individual scientists who are often the first to notice or measure the impacts. As awareness is raised, the interest generated gives rise to informal consensus-building 'networks', which discuss ways of harmonising activities such as data collection and exchange. Often such networks depend on the resources of a key institution or on external support to survive. Finally, the networks evolve into centrally organised, self-supporting bodies, which are recognised or even adopted by government.

Not all information responses occur in this way; some may be directly initiated by governments from the beginning, or indirectly via externally sponsored projects. By whatever means the profile of biodiversity information is raised, its impact on decision-making will be determined by the extent to which it is *relevant* to decision-making needs and, in the case of governments, relevant to immediate *policy* considerations.

Making the provision of policy-relevant information a clear aim of the awareness-raising process helps to provide a focus to collaborative activities. Information can be explicitly developed to reveal short- and long-term impacts on biodiversity, and suggest ways in which policies can be changed to ease the problems highlighted.

Biodiversity is a multi-scale, multi-disciplinary issue, and thought must be given to ways in which stakeholders can develop information co-operatively as opposed to pursuing only their own interests. Only co-operative action can mobilise the wide range of expertise characteristic of biodiversity issues.

1.2 INFORMATION NEEDS

With limited resources available to produce information, the setting of priorities is critical. To ensure policy-relevance, priorities should reflect the *information needs* of decision-making groups. These may not be clearly articulated by the group concerned (who may have only a hazy idea of their requirements), but do have to be agreed by the majority of stakeholders producing the information (see Chapter 2 for a discussion of organisational structure).

There are many issues in biodiversity which require information. Most result from direct physical, chemical or biological pressures exerted on the environment by human economic and development activities. It is useful to consider just a few of these to illustrate their breadth:

- Habitat and landscape conversion (eg of forests to agriculture), and its effects on human welfare.

  Information needs: the economic and social benefits of ecosystem and landscape protection, eg sustainable use revenues and environmental services, including their distribution amongst human populations; the key pressures applied to landscapes by human activities, and the resulting trends in landscape condition; the current policy and legislative framework for conservation (including protected areas), and costs of associated programmes and projects (including opportunity costs of development).

- Decline in commercially or ecologically significant species

  Information needs: the economic and social benefits of species or groups (eg for food, raw materials, medicines, tourism); additional, ecosystem-related benefits (eg keystone species); the distribution and status of wild populations, including current trends and potential for extinction; the key pressures facing species (eg habitat conversion, over-exploitation, invasion of exotic species); the quality and extent of protective legislation; the costs and purpose of species-related conservation programmes and projects, including ex-situ measures.

- Erosion of genetic resources (eg wild ancestors of domestic breeds or cultivars)

  Information needs: the economic benefits of indigenous genetic resources (eg food security, biotechnology potential); the distribution and status of selected genes; the driving forces of genetic erosion (in addition to those impacting landscapes and species); the quality and extent of protective legislation (eg on import of new varieties); the costs and purpose of protection programmes, including ex-situ measures such as gene banks.

Many other issues could be described, including the loss of indigenous knowledge of traditional uses and values of biological resources, the impact of global climate change, plus sector-specific issues relating to sustainable management practices.

Different issues take precedence according to the particular pressures exerted on biological resources in the locations concerned, and the extent of public, media, government and other interest in what is happening. Issues also change in time, sometimes very rapidly; they arise, come to the attention of decision-making groups, and then disappear - perhaps to resurface in another form at a later date. The key to effective use of information is knowing when and to whom information should be delivered.

1.3 THE POWER OF INFORMATION

Information empowers its audience by:

- providing a range of options
- providing a wider context within which to assess impacts and options
- adding to a common basis of agreed facts on which to base debate
- discouraging options with predictably adverse consequences.

In general, the audiences we are most interested in influencing are the senior managers in government, NGOs and the commercial sector. However, the indirect influence of the public, media, community groups and international bodies and conventions should also be recognised.

Audiences at all levels have little time to interpret raw, unprocessed data. They require information which can be quickly and easily digested, yet significant and lasting in impact (see Chapter 6 for a full discussion). To fulfil this objective information should be:

- available when the 'window of opportunity' for decision-making arises (ie timely)
- easily and quickly understood (eg presented using single numbers, trends, maps or charts)
- relevant to immediate policy needs
- delivered by recognised channels into the decision-making process
- based on sound scientific principles
- accessible in standard formats or interfaces which require minimal prior knowledge to use
- available at minimal cost in terms of time, money and administrative overheads
- free from unnecessary restrictions on use
- accompanied by full acknowledgement of intermediate products, data sources and intellectual property (an 'audit trail').

These characteristics form the basis of a group of information products known as *indicators*, which are time-varying measures of policy performance, accepted as reliable by many sectors of society. Good examples from the financial domain include the Dow Jones and FTSE indices from Wall Street (New York) and the City of London respectively. The frequently used GNP and GDP figures estimating nations' economic performance are also examples. In the case of biodiversity, the development and use of specific indicators will enable governments and other groups to measure progress towards sustainability targets in an open, objective manner (for a detailed discussion of environmental indicators see Hammond *et al* 1995).

In some cases, the strategic release of information can help define environmental agendas. A good example is the release in 1990 of global 'greenhouse' gas emissions by the World Resources Institute (WRI/UNDP 1990). All major countries were ranked according to their level of emissions, causing immediate attention to and rapid alteration of policy in many cases. Equivalent impacts are possible at the national level.

The above example illustrates 'decision-making by disclosure' - ie the delivery of information to policy-makers via the public domain, rather than by more traditional channels. This may be appropriate in some situations but can be counter-productive in others. An understanding of the political, social and legislative climate of the country is required before deciding how information should be released to maximum effect.

1.4 INFORMATION AS A TOOL

In the same way that labour, transport and buildings enable managers to run their businesses more efficiently, so does information. But like these other production factors, too much information is costly and unnecessary. The key to effective use of information is to focus on essential information only - i.e. that which is needed to set and achieve policy goals. Further, when several organisations join forces to develop information, costs can be cut and efforts synergised to develop products beyond the capabilities of individual agencies.

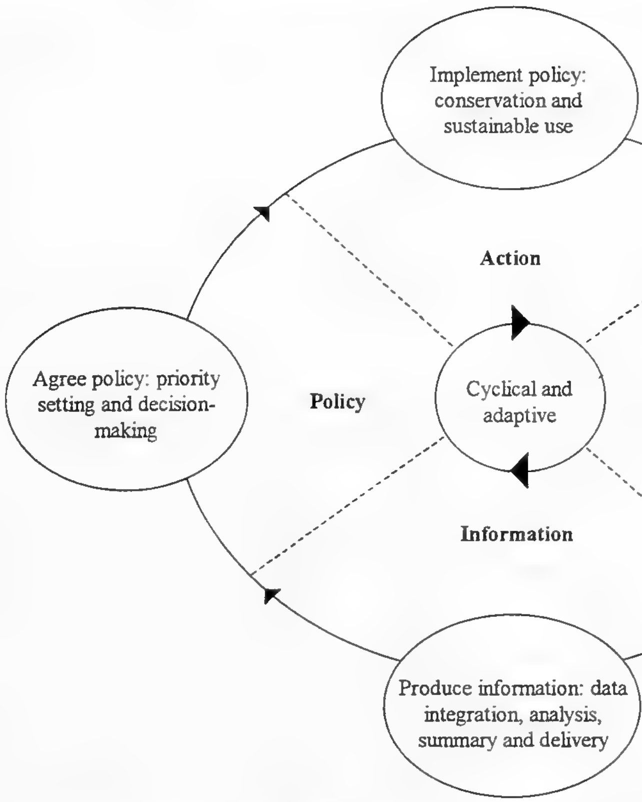
So how can the use of information in an organisation increase productivity? The operation of any kind of business can be represented in terms of its constituent processes and information flow (see Section 3.3.4). When current operations are examined it is often possible to detect information blockages, 'black holes', gaps, and overloads, which hold back the productivity of individual staff and the business as a whole. For instance, how can a local resource manager plan sustainable extraction regimes without knowing the regeneration potential of the resource?

The analysis of organisations in terms of information supply, demand and usage is helpful in identifying priority areas for investment. Attention to information management practices can improve overall corporate efficiency and increase the capacity to deliver information to others - thereby earning credibility.

To see where information fits into the business of conservation and sustainable use, Figure 1 illustrates a typical 'management loop' containing four processes: policy, action, data, and information. These may be addressed concurrently or sequentially, and may be revisited at any time. Two features of the loop are that it is *cyclical* and *adaptive*, achieving policy objectives in a progressive manner (see Section 5.2).

Although Figure 1 is similar to previous models for national biodiversity planning (see for example UNEP 1993 or WRI/UNEP/IUCN 1995), the role of information is made more explicit. Monitoring the effects of management actions - or lack of actions - on biodiversity, is clearly linked with data collection and management; and the provision of information for policy review ('closing the loop') is linked with data integration, analysis and delivery.

Figure 1: Management loop for conservation and sustainable use



Two processes should be undertaken before entering the loop:

1. Agreement of the key *issues* in biodiversity conservation and sustainable use, by means of an initial (possibly rough) assessment of the social, economic and ecological objectives.
2. Agreement of the roles and responsibilities of different agencies in establishing and co-ordinating information and monitoring programmes.

The transition from *exploitation* of biodiversity to *sustainable use* will require intensive information and monitoring. Biodiversity professionals must respond by developing information systems that serve policy-making needs, a goal which depends on the participation of a wide variety of individuals and agencies.

2

Information Systems - The Framework

2.1 INTRODUCTION

Information provides essential support for an organisation's corporate goals, whether that organisation be a village, a resource management agency, a nation or a multilateral bank. 'Market' intelligence applies just as much to environmental information as it does to economic or political information. Without proper management and use of environmental information, a village can go hungry, an agency or nation can degrade or destroy valuable resources and an international institution can oversee programs that have impacts opposite to those intended.

Information, unless continuously maintained and upgraded, degrades in quality and value. Data, the raw material for information, must be regularly gathered, managed and processed into useful information if an organisation is to achieve its core objectives. Under present circumstances, gaining access to crucial data can be difficult and expensive, being often frustrated by political, organisational and even personal barriers.

A key challenge to governments and other entities is to minimise these barriers, to 'reduce the transaction costs' of using data and information in pursuit of environmental sustainability and other desirable ends. Freeing up the flow of data confers distinct advantages on resource management and policy agencies at all levels. It enhances effectiveness and generates new 'business', whatever their area of activity. It also enables agencies to combine their information resources, generating totally new products that increase their collective impact on decision-making processes.

The people, data, processes and tools necessary to achieve these impacts are referred to, collectively, as an information system or, in cases where multiple agencies are involved, an information network.

The primary ingredients of an information system are described below:

- People

  These are the stakeholders of the information system, whatever their function. This is an all embracing concept which includes:

  - ✓ the people who originally collect data (agriculturalists, biologists, ecologists, economists, indigenous peoples)
  - ✓ the people who develop and maintain the information system (systems analysts, designers, technical support)
  - ✓ the people who create and disseminate information (data analysts, publishers)
  - ✓ the people who manage the process
  - ✓ the people who receive and are empowered by the information (decision-makers, general public, international community - these groups may overlap with previous groups).

- Data

  Data are the core of an information system and occur in a variety of types, formats and media, originating from one or many agencies. Examples are paper maps and reports, computerised specimen records, and air pollution values.

- Tools

  These include filing cabinets, box-files, record keeping books, computers, data input devices such as scanners, output devices such as printers and plotters, general purpose software, data management software, and specialised data analysis and publishing software.

- Processes

  Processes define what the people do with the tools in order to manage and interpret data effectively and efficiently to achieve the desired information products.

Modern information systems make extensive use of computers, but this is not essential; the same principles are retained whether or not computers are applied, such as the need to structure data efficiently, the need for data to flow between different processes without restriction, the need to integrate data, and the need to create simple, interpretable products. Many of these processes are highly specialised and require human intervention.

2.2 DESIGN CONCEPTS

2.2.1 Overview

It is tempting to see a multi-agency information system as an opportunity to centralise a wide range of data resources in a single, possibly new, location. Whilst this may be efficient within a single agency - where individual feelings of data ownership are subsumed by mainstream corporate objectives - it is impractical in multi-agency situations. Most agencies expect to retain full rights over their data when participating in collaborative projects, including the right to manage data at their own premises.

The key to effective data management is to have each dataset managed by the agency best qualified to ensure its quality and accessibility. The concept of 'custodianship' provides a useful means by which such agencies can be identified, and the associated rights and duties disbursed. These include the responsibility for collection, management, and documentation of the dataset, and for determining the conditions under which it can be accessed and used.

The key to effective data management is to have each theme, dataset or entity managed by the agency, group or individual best able to ensure its quality and accessibility.

2.2.2 Custodianship

Custodianship is a generic concept which may be applied at all management levels. Every dataset - and this is especially true of nationally significant datasets - should have one and only one custodian. The concept is very practical: custodianship encourages a sense of 'ownership' of data, which contributes to their quality.

At the national level, responsibility for data *themes* is usually allocated among government departments. For example, land infrastructure such as administrative boundaries, topography, settlements, roads and rivers might be assigned to a department of survey and mapping. At the agency level, responsibility for specific *datasets* may be allocated to sub-departments, units, or other recognised groups. Similarly, within such groups individuals assume responsibility for maintenance and development of sub-components, or *entities,* of a dataset.

A distinction should be drawn between data themes and datasets (Busby 1994). A theme, such as topography, can consist of a large number of diverse datasets. Responsibility for the theme could be allocated, for administrative reasons, to one specific agency. That agency may then assume custodianship for one or more topographic datasets. However, such an administrative arrangement must not prevent other agencies from developing topographic datasets to meet their own requirements - and for which they would wish to become custodians. A good example is the defence forces wishing to develop a vegetation dataset to enable them to plan heavy vehicle exercises. The attributes needed for that task would be different from almost all other agencies. Thus they would develop and manage that dataset and, assuming that security was not an issue, make it accessible to other agencies. Custodianship, therefore, applies at the dataset level; it should not be applied at the data theme level.

It is accepted that environmental data are not easily categorised and overlap in jurisdiction can easily occur. The way forward is to designate one agency the overall custodian of a dataset, and allow other agencies to manage sub-components (*entities*) of the dataset. An example would be a species dataset held in a protected areas management agency. Data on the distribution and economic value of the species are held by the protected areas agency, but the list of names used to reference the species may be managed by a more specialist custodian such as the local museum or herbarium.

The most appropriate agency to manage a dataset is likely to meet one or more of the following criteria:

- has sole statutory responsibility for capture and maintenance of the data
- is the first to record changes to the data
- is the most competent to capture and/or maintain those data
- has the confidence of users that it will continue to meet its commitments to data collection and maintenance.

In accepting the custodianship of a dataset, the following responsibilities are assumed:

- define and maintain quality standards
- keep the dataset up to date
- ensure the continued integrity of the dataset
- ensure appropriate access to the dataset
- maintain documentation on the dataset
- advise on appropriate uses of the dataset.

> *Each dataset should have an identified custodian to manage its development, quality, and external access.*

2.2.3 Architecture

The concept of custodianship suggests that, unless exceptional circumstances prevail (eg the capacity to manage data at a certain agency is inadequate), the architecture of a multi-agency information system should reflect the rights of agencies to manage data at their own premises. In practice this means that a 'distributed', rather than 'centralised', information system architecture is required (see Figure 2). Centralised architectures are useful in more tightly controlled situations where, for reasons of urgency or security, data are relocated from their owner for management elsewhere.

Distributed, or network architectures have the advantage that the development of the information system occurs in all the collaborating agencies, rather than in only one, centralised location. As a result, the benefits of collaboration are gained by many participants in the project. The distributed approach also fosters ties between agencies, leading to mutual improvements in security and performance (see Section 2.3.2).

> *To respect the rights of custodians to manage their own data, information systems should be designed with distributed architectures.*

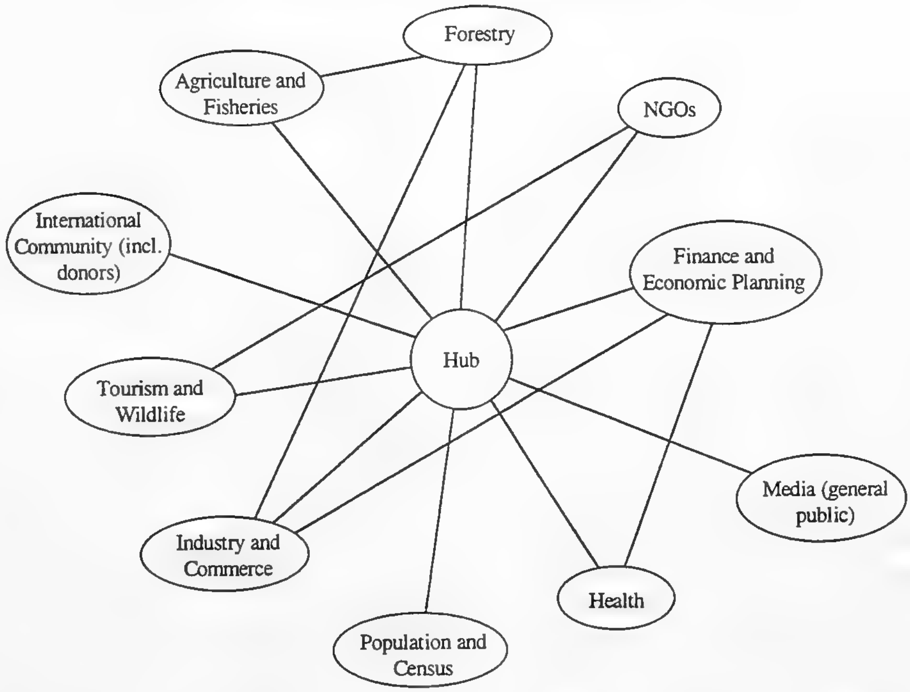
2.3 NETWORK CO-ORDINATION

2.3.1 Overview

The greatest challenge with distributed data management is network (inter-agency) co-ordination. Some unit, team, or other group must take responsibility for facilitating joint action by the agencies involved if the full benefits of collaboration are to be achieved. This group, which lies at the centre or 'hub' of the network of participating agencies and users, has the following essential functions:

- influence decision-making in a timely and authoritative manner, via the delivery of information (preferably in the form of easy to interpret *indicators*) to decision-makers in the public and private sectors, the media, NGOs, and international community
- promote dialogue between agencies in the form of meetings, workshops, correspondence, newsletters, and other forms of information exchange

Figure 2: Multi-agency information system or 'network'



- define specific information objectives, such as the agreement of standards for data collection and reporting, and the formation of integrated information products (these should be *synergistic*, ie greater than the sum of what could be achieved by individual users)
- work with each partner to assess their strengths and weaknesses, and arrange capacity building to improve data management practices
- liaise with development assistance organisations to gain support for key objectives.

Figure 2 illustrates the position of the hub within a typical network consisting of a wide constituency of users, ranging from sectoral agencies in government, to NGOs, the media (general public), and the international community (eg conventions, donors, global data centres). Note that users are free to maintain one-to-one linkages with other users, in addition to their link with the hub.

To operate most effectively, multi-agency information systems should be co-ordinated via a central hub.

2.3.2 Data Exchange

The issue of data exchange is frequently raised when groups of agencies meet to discuss collaborative information projects. The potential for infringement of intellectual property, abuse of copyright, or inappropriate application of data, are legitimate fears which tend to impede progress in this area. As a result, data exchange agreements are perceived to be difficult to negotiate.

The concept of custodianship can be called upon to resolve this problem by providing an umbrella under which agencies make their data available to each other. In particular, agencies should understand that data exchange:

- is mutually beneficial to the recipient and provider - the value to the recipient comes with use; the value to the provider comes with credibility for being of service (paving the way for future exchanges and access to value-added products)
- fosters an atmosphere of mutual trust and co-operation between data management agencies, adding to their long-term security
- does not require agencies to give up their legitimate rights as custodians, including their responsibility for data collection, their right to update and manage data as they see fit, and their right to specify how the data should or should not be used (eg data may be used for government planning or research but not for commercial purposes)
- can be regulated to ensure that copyright, intellectual property and other legitimate rights are protected.

Individually, or as a group, agencies should develop simple operational procedures for data exchange which minimise administrative, cost and other restrictions on use (this process is elaborated in Section 5.4.2). These may take the form of 'Memoranda of Understanding' linking two or more named institutions, or generic protocols which may be applied in all circumstances.

Data exchange is beneficial to both recipient and provider; it fosters an atmosphere of mutual trust amongst agencies; it does not require custodians to give up their legitimate rights; and it can be regulated by simple operational procedures.

2.3.3 Legal Considerations

In many countries there is considerable ignorance over the law concerning the rights and obligations of originators and compilers of biological data. For instance, there appear to be no explicit or binding obligations under present UK or European legislation, or through international agreements, which require any individual or agency to maintain biological data (Burnett *et al* 1995). The need is implied, however, in numerous international agreements and initiatives including the CBD.

At the international level, the exchange of information on biological resources may impinge on legal and conceptual views of sovereignty and security, particularly where the information concerns government policies and legislation. Despite being keen to promote the flow of information on biodiversity issues, the CBD is also very conscious of this issue. It is clear that concerns over the misuse of information for strategic or political purposes must be addressed before the desired level of information exchange will be achieved.

The exchange of certain kinds of information, particularly on biotechnology and other 'enabling' technologies, is often subject to national and international copyright and patent law. The precise details, including penalties for infringement, vary greatly according to the nature of the information exchanged, and the legal establishment of the country concerned.

In general, copyright affords protection to a biological dataset in its permanent form, independent of how the data are disseminated to others (eg in writing, illustration, broadcast etc). The originator can assign or license copyright to another individual or agency, provided agreement is made in writing. However, the 'moral rights' (such as the right to be acknowledged in publications and the right not to allow unauthorised alteration or misrepresentation) remain with the originator. Thus an individual or agency wishing to compile or change a biological dataset must have written permission from the originators if they are not the owners or originators themselves (Burnett *et al* 1995).

Data providers are also subject to certain liabilities. In the event of incorrect data being provided, or harm caused, liability could fall on the originator of the data, its present custodian, a third party agency which has provided the data or on all of these. The situation is most serious when 'negligence' is detected - for instance no reasonable attempt to ensure data quality was made, or data corruption resulted from poor operational practices (see Section 5.3).

Iinformation should be kept confidential if its provider has not consented to its release. This may occur in cases where information might compromise the survival of a species or increase the risk to a landscape.

To reduce the risk of liability in the event of incorrect data being released, high standards of data quality must be maintained.

2.4 ORGANISATIONAL STRUCTURE

2.4.1 Overview

The success of the co-ordinating 'hub' (see Figure 2) may be judged by the degree to which stakeholders feel involved and responsible for overall project management, and the extent to which collaborative objectives are achieved (see Section 2.3).

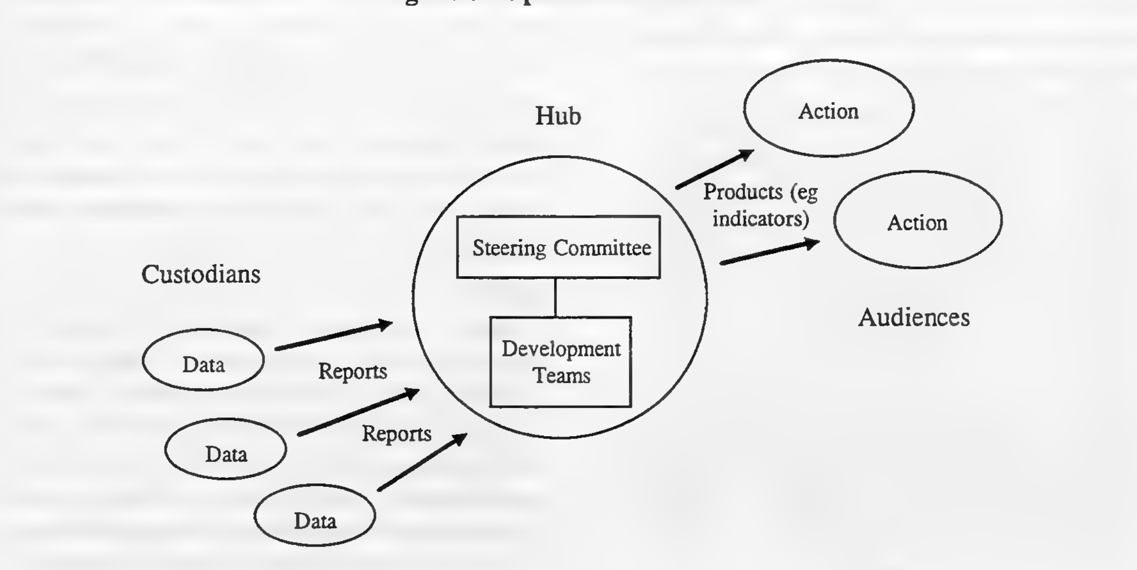
A commonly used structure for implementing the hub comprises the following two bodies:

- Steering Committee

  This is a high-level management group representing key institutional stakeholders in the information system. The Steering Committee provides leadership and authority throughout the project lifetime, and 'signs-off' after the completion of each major development or product. It is responsible for selecting and managing Development Teams (see below), and for bringing forward solutions to other collaborative objectives.

  The Steering Committee, which may be

Figure 3: Operation of the hub



composed of respected and influential members, may continue to exist after the completion of the physical information system, as a leading policy group on environmental information.

- Development Team(s)

  Under the umbrella of the Steering Committee, Development Teams are responsible for developing the capacity of custodian agencies to manage their data effectively, and for specifying procedures for summarising data and producing information products. Expertise may be required in the areas of strategy development (including user needs assessment) (Section 5.5.4), information production (Section 5.5.3), and technical support (Section 5.5.2).

  Development Teams should be drawn from the existing human resources of custodians where possible, supplemented by contracted experts to cover the required skills. After completion of the project, members of the Teams may be retained for technical support, training, and related capacity building activities.

The purpose of this two-tier arrangement is to separate decision-making processes which involve political and organisational issues, such as resource allocation, transparency, jurisdiction over data and services, and legal implications of data exchange, from operational processes, which concern the activities of teams charged with building components of the system and information products. The latter must be free to explore technical issues in an atmosphere of trust and confidence, free from unnecessary burdens imposed by higher-level processes.

Multi-agency information systems should attempt to build an organisational structure consisting of a high-level Steering Committee and one or more expertly staffed Development Teams.

2.4.2 Information Flow

Lying at the centre of the information system, the hub is in a unique position to facilitate information flow. This involves harnessing the potential of custodian agencies to produce collaborative information (see Chapter 6).

Figure 3 illustrates one component of this process in action. Three key activities should be observed:

1. data are summarised in standard reporting formats by custodian agencies, and sent to the hub on a periodic basis
2. reports are integrated, analysed and summarised by Development Teams at the hub, and packaged into information products
3. the Steering Committee approves information products for release, and decides when and to whom they are delivered.

Unless the hub agency is also the custodian of one or more datasets, it does not *need* to manage any such data itself (individual custodian agencies perform this duty). However, responsibility for one theme should lie with the hub in order to assist its co-ordinating role: a register of contacts, capacities and metadata able to support biodiversity data management and planning.

As well as containing full contact details (name, address, telephone etc.) of the individuals, groups and agencies concerned, details of their particular expertise and resources (eg data) should also be recorded.

Contacts may range from local community leaders able to document indigenous knowledge or implement conservation regimes, to a wide variety of national groups in the research, planning and resource management sectors, donor agencies, international NGOs and other global organisations. The objective is to be able to match identified needs with appropriate support, and place overlapping activities in touch with each other to share experiences and data. Much of the required information can be obtained via an 'institutional survey', which is described in an accompanying document, *Guidelines for National Institutional Survey*.

To match needs with appropriate support, the hub should maintain registers of contacts, capacities and data resources.

2.5 PRIORITY SETTING

2.5.1 Overview

Once the basic architecture of the information system has been defined, and an organisational structure has been formed to develop and manage it, the next challenge is to design a strategy for implementation.

In order to establish priorities within the strategy (to ensure that processes are undertaken in the appropriate order), it is useful to maintain a mental picture of the overall development process. One possible 'framework' is illustrated in Figure 4.

The steps outlined below, which mirror the framework illustrated in Figure 4, are intended to help prioritise information system building activities.

In cases where particular processes have already been accomplished (eg an information needs assessment), the framework may still be useful in suggesting next steps, or drawing attention to missed, or under-emphasised activities.

It should also be noted that the framework is not rigid or prescriptive; flexibility is likely in the ordering of processes due to some work having already been undertaken, the need to revisit earlier processes, and a variety of other local constraints and needs. However, in the interests if simplicity, the many feedback loops which connect the processes together have been omitted.

2.5.2 Key Steps

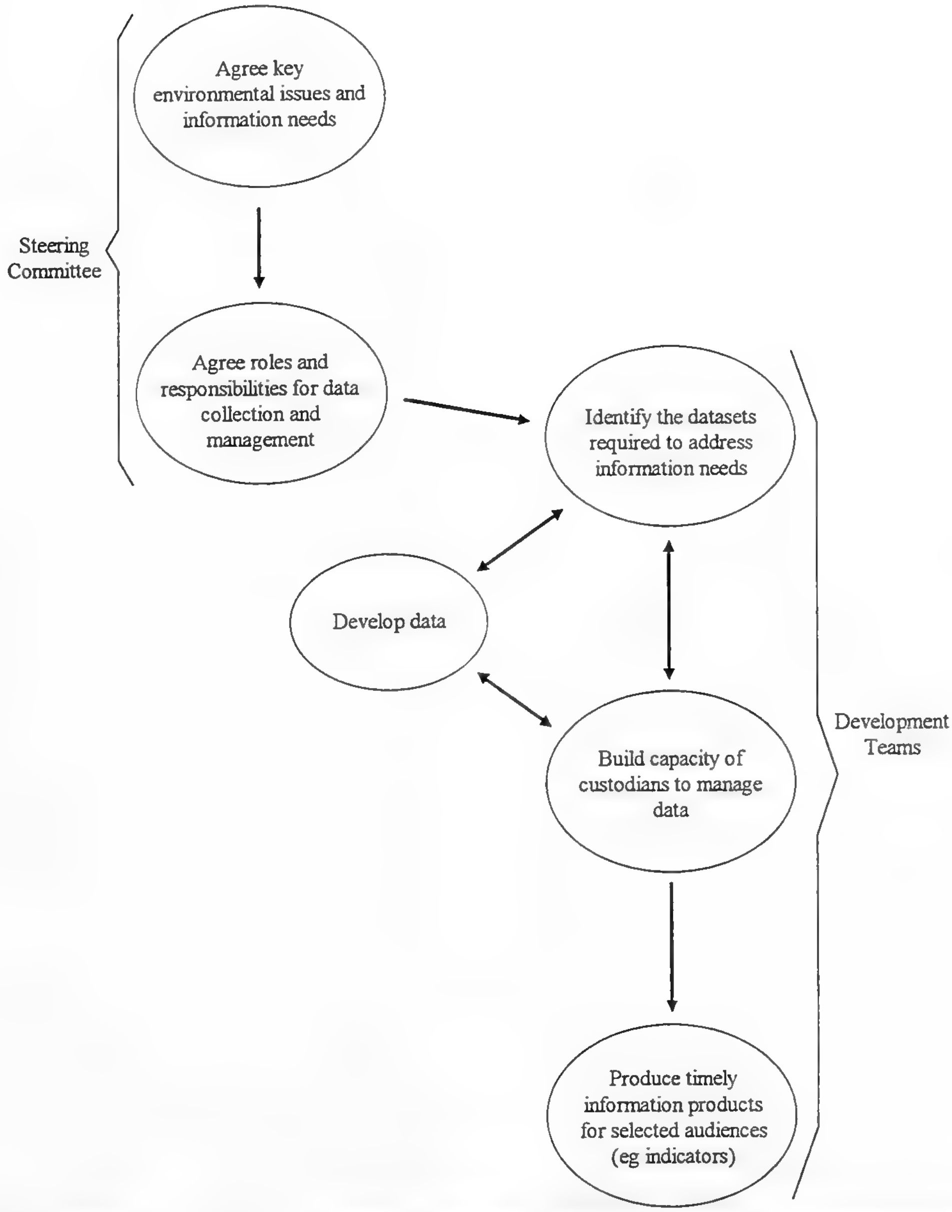
1. Agree key environmental issues and information needs (see Chapter 1). In this step the Steering Committee, or other body established to facilitate the development of the information system, decides which issues are most urgently in need of information support (hence action). This is a highly consultative process, requiring broad participation from all stakeholders in the information system, including some who might not be directly involved (eg high-level decision-making groups). During this step goals are set in the form of key products that the information system will provide, for example simple-to-interpret indicators of environmental pressure and change.

2. Agree roles and responsibilities for data collection and management (see Section 2.2.2). In this step, the Steering Committee agrees the custodians of key data themes and vital services (eg lobbying, brokering). This with the identification of issues and needs, this is a highly consultative process. The aim is to assign broad areas of custodianship and pinpoint themes which are not well covered (see Step 4), so that appropriate linkages with other agencies can be established.

3. Identify the datasets required to address the information needs (see Chapter 4). In this step the data resources of the custodians are analysed, and the specific datasets required for product building are identified (see below). This process is technical rather than organisational, and should be undertaken by a qualified Development Team.

4. Develop data. It may be that previous steps expose gaps or omissions in the collective data resources of custodian agencies. These may have gone unnoticed prior to the information system project,

Figure 4: Framework for information system development



since they may be 'collective' gaps, rather than gaps attributable to a particular institution. The task of the Steering Committee is to enable custodians to develop their data by facilitating 'gap-filling' exercises where critical data are absent or insufficient, or quality improvement activities where data need revising, improved management, extension, or repair. The work is undertaken by Development Teams formed within the agencies concerned.

5. Build capacity of custodians to manage data (see Chapters 3, 4 and 5). This step, which involves building the capacity of custodian agencies to manage key datasets effectively, is undertaken by one or more Development Teams formed by the Steering Committee. Key activities include user needs assessment, system design, development, implementation and operation.

6. Produce timely information products for selected audiences (see Chapter 6). This step involves establishing the necessary procedures to enable custodians to report their data to the information systems hub, where they are integrated, packaged and communicated to target audiences. The latter requires proper attention to the physical impact of the message (ie simplicity, length, use of colour, graphics etc); the timeliness of its release (ie judging a 'window of opportunity'); and its method of release (eg to a government minister, press conference, scientific conference, workshop, informal grouping, or inter-personal dialogue). Computing facilities at the hub may need to be established to enable it to process reports from custodians effectively.

3

Information Systems Development

3.1 INTRODUCTION

The concept of custodianship implies that data should be collected and managed by the agency which is most appropriate and best equipped to do so. Since biodiversity is a very wide ranging topic, spanning many agencies and disciplines, the application of custodianship results in a distributed information system consisting of loosely linked datasets managed in separate locations (see Figure 2).

To operate effectively, three fundamental activities are necessary in such a system:

- Regular collection (monitoring)

  Many agencies are good at data collection. However, the time dimension may be lacking from their data - the paradigm to use is monitoring, not collection - and the techniques used may not be consistent over time or with other agencies. To reveal environmental trends, data should be collected in standard formats, via standard techniques, and over long periods of time. One-off studies may be interesting for many reasons, but consistency of results is almost certainly more useful in the long-term.

- Management and accessibility

  For information to flow between agencies, and from agencies to other audiences, data should be managed in ways which promote accessibility. Various principles and techniques are necessary to achieve this, including the design and development of local information systems (this chapter) and, more specifically, computer databases (see Chapter 4). It is the task of all individual stakeholders in the network and, in particular, its hub to ensure that sufficient resources and expertise are mobilised to develop the capacity of agencies to manage data effectively.

- Summary into information

  Although the architecture of a multi-agency information system is distributed, some co-ordinated activity is necessary to firstly summarise the data collected by custodians and secondly build collaborative information products (see Section 2.3). This activity is facilitated by the network hub, which maintains contact with all the custodians and monitors the status of their data (see Figure 3).

Within the context of the overall information system, partner agencies have two goals: to develop their own data for improved corporate productivity; and to integrate their data with other agencies to achieve results beyond individual capacities. Thus improvements in data management capacity are immediately beneficial to the agency concerned, as well as the wider network.

Depending on the profile given to information within an agency, and the resources which are available, projects may already be underway to increase information usage, boost data management effectiveness, and even implement localised information systems. The potential for collaborating with or building on existing work should be investigated by all agencies before embarking on new projects, since the experiences gained may be extremely valuable. However, in many situations custodians will wish to initiate new projects to manage their data, and in such cases assistance may be required from the network hub.

In order to address the needs of different custodians, a variety of approaches to information system development may be required, and these should always be debated in an open, consultative manner (see Section 3.3). For example, a system set up to manage a plant genetic resources dataset in a ministry of agriculture may differ greatly from a system set up to manage sustainability indicators in a forest department. It would not be efficient to merge the two datasets together, nor would it be acceptable to the agencies concerned.

Although every information system project will have its own objectives, generic methods of project management are useful in structuring the design and development process.

The remainder of this chapter considers such techniques - referred to as system development methodologies - whilst Chapter 4 focuses in greater depth on the specific issue of database design.

3.2 HISTORICAL CONTEXT

As the use of computers in information management has expanded, methodologies for the development of information systems have steadily matured. These originally arose to address the problem of excessive cost (resources and time), which often exceeded original estimates.

In the late 1960s and early '70s, a standard project life cycle became accepted as a means of structuring information system projects. Given the constraints of the technology at that time (for example mainframe architectures, punch card processing, and languages such as FORTRAN and COBOL), projects tended to be managed by computer specialists. When eventually delivered, the systems were subjected to criticisms such as 'not what I wanted', 'incomplete' and 'unworkable'. Two factors contributed to this:

- long development periods during which users altered their requirements
- difficulties in phrasing user needs in complete and unambiguous ways.

In the early 1970s the first of these challenges was partly addressed via concepts such as structured programming and structured analysis. The tools which were developed to support these techniques prepared the ground for many of the development tools we use today - tools which relieve much of the burden of programming. The driving force was productivity, since the cost of human resources was a key consideration in the overall project.

A class of system development methodologies grew up around the structured programming paradigm. Collectively, these are referred to as the Structured Development Life Cycle approach, in which development is carried out in a series of structured phases. Different variants of the life cycle are advocated by different countries and organisations, some of which are accepted as 'standards' in industry and government.

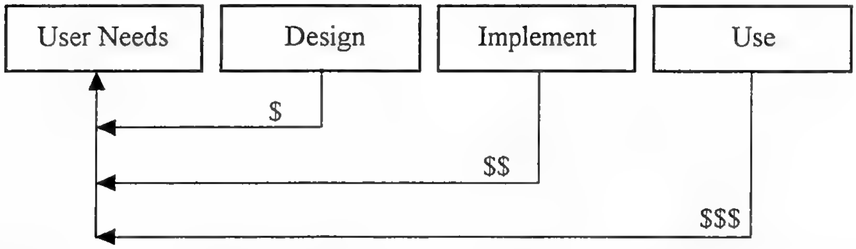
Computer performance has increased markedly since the 1970s. Project development is now centred around the 'desktop', with powerful, sometimes graphical languages being introduced for accelerated system design. With this revolution came the introduction of 'prototyping' tools capable of modelling the finished product quickly to generate feedback from prospective users. This led the introduction of new development methodologies in which prototype systems are modified on the basis of feedback from prospective users.

3.3 USER NEEDS ASSESSMENT

3.3.1 Overview

When building information systems, the earlier that problems are identified the easier (and therefore cheaper) it is to correct them. Correcting an error during the design stage is normally a simple, paperwork task; correcting an error after a design has been translated into a working system is more costly (this may require equipment modifications); correcting an error after

Figure 5: Relative cost of change during information system development



users have begun to employ the system for their work is more costly still, and may involve retraining of staff in addition to equipment modifications (see Figure 5).

It is important to assess user needs rigorously before embarking on an information system development, and refer to these needs often as work progresses. Without proper attention to user needs assessment, time and money can be wasted on systems which are not cost-effective (eg fail to deliver the required products), leading to dissatisfaction and eventually loss of confidence in the project by stakeholders.

The key challenges in user needs assessment are therefore:

- to reduce unnecessary costs and delays during system development
- to promote ownership of the development process by stakeholders.

Clearly, the solution to these challenges will be different in each project. However, in most cases the principle objectives of a user needs assessment can be summarised as follows:

- to define the users of the information system, especially the audience for its information products
- to determine the priority information needs of these users
- to set objectives for information system performance, including defined products and services
- to establish participative, collaborative approaches to information production and use.

The main product of the assessment is a document known as the 'functional specification' of the information system. This describes the background to the information system project including the justification, cost-benefit analysis, description of key stakeholders (including their capabilities and needs), and definition of products and services.

The specification also comprises technical details such as an inventory of essential datasets, diagrams illustrating information flow between system processes, and definitions of the major database structures. This is done at a formal, conceptual level since the specification is quite independent of equipment issues (eg hardware or software); indeed, it should be free from any kind of implementation details.

The size and formality of the specification will vary according to the complexity of the system proposed. For instance, a complex project involving several sectors and agencies might be broken down into a series of sub-projects, each with their own functional specification.

An indication of the importance of the user needs assessment is provided by Richardson (1994), who claims that this step "took 80% of the time of the start-up phase" of the Environmental Resources Information Network (ERIN) information system in Australia, and that "great self-control was needed not to be 'busy' purchasing hardware, software, and data until these matters were settled".

Most standard text books on information systems development devote a chapter to user needs assessment, as do more specific books on GIS implementation. Two examples are Powers and Cheney (1990) and Aronoff (1989). A useful guide to establishing needs for GIS can also be found in Wiggins and French (1992) and guidelines for the requirements phase for general information systems development in the *Model Software Development Standard*.

To reduce unnecessary costs and delays during system development, emphasis should be placed on the early stages of the development process, particularly user needs assessment.

3.3.2 Initial Steps

Active consultation is essential during a user needs assessment to promote participation in the development process and reveal needs which cannot be 'guessed' reliably by developers. Conversely, consultation allows developers to explain the potential applications and limitations of information technology to different users.

Assessment projects often begin with a workshop attended by representatives of the major stakeholders and technical experts who will contribute to the information system. This workshop should attempt to reach agreement on:

- which environmental issues are the highest priority
- what information is required to support decisions on these issues (content)
- what long-term information and monitoring programmes are required, and who is responsible for implementing them
- which audiences require information most urgently, and how best to reach them (delivery)
- how and when information should be presented (format and timeliness)
- which data collection and management standards will be followed
- what mechanisms are required for data exchange and cost recovery (eg 'Memoranda of Understanding')
- what are the main capacity building needs (eg technical and human resources).

More detailed consultations between stakeholders and members of the Development Team will be necessary as the assessment progresses. These usually take the form of questionnaires, interviews, brainstorming sessions and working groups (see Section 3.3.5), during which stakeholders are invited to outline institutional strengths and capacity building needs, and suggest specific collaborative objectives. In response, representatives of the Development Team may probe the operational procedures of the user's organisation to judge how best to implement requests. Multiple consultations may be required to deal with operational issues such as data availability, quality assurance, operating procedures and data security.

In large projects, formal techniques such as data modelling (which results in entity relationship diagrams), process modelling and prototyping are used to structure the assessment results. An example of a formal specification (for BirdLife International) can be found in Van Dijkhuizen (1994), and a less formal example (for the UNEP Office of

Harmonization of Environmental Information) in Crain (1992).

To determine the overall issues, objectives and challenges of the user needs assessment, it is constructive to hold an initial workshop attended by major stakeholders and experts.

3.3.3 Data Needs

Having decided the key environmental issues and information needs, the task of the Development Team is to determine which datasets are required to support them. For instance, the need "to be able to decide on enhancements to a national parks system", may require data on the current extent and status of protected areas. Similarly, the need "to decide whether to permit bio-exploration" in a certain region, may require information on the ecology, biodiversity, traditional uses, and cultural values of the region.

Data modelling is commonly used to facilitate the transformation of information needs into data requirements (see Section 3.3.5). In this technique, primary data 'entities' are depicted graphically, and their relationships to one other made explicit. This is useful for communicating the nature and structure of perceived data requirements back to users for verification, and also serves to consolidate ideas. At this stage data modelling should be restricted to high levels of generality, and make no reference to where or how the data will be obtained or managed. More detailed data modelling takes place during information system design (see Section 4.5).

Currently available datasets should be catalogued on paper or electronically in a metadatabase (see Section 5.3). This enables gaps to be determined by comparing existing datasets with those which are required. Data needs can then be expressed in terms of existing datasets (which may need enhancement) and new datasets which are required to cover gaps. During this process it should be remembered that gaps should only be filled if there is sufficient justification for doing so. Data collection should always be linked to the development of priority information products, rather than being treated as an end in itself.

The assessment of data needs should lead to the following outputs:

- table of required datasets, indicating content, current custodianship, access method, data type (eg tabular, text, spatial, graphics), and quality estimate
- generalised data model
- preliminary data dictionary.

3.3.4 Processing Needs

Various processing tasks are necessary to transform data into information products. These should be documented to enable appropriate facilities to be built into the information system. Typical processing tasks include data integration, analysis, validity checking, updating, and reporting. It is convenient to divide data processing needs into three categories: management, analysis and production.

- Management processes ensure that data are maintained securely and made available for widespread use. Typical processes include dataset documentation, quality assurance (error detection, update, backup), application of standards, database development, and negotiation of data exchange agreements. Associated processing needs are those which facilitate the use of data within an organisation, such as file conversion and exchange, procurement, messaging (eg electronic mail and other on-line services), and technical support and training.
- Analysis processes are applied to one or more actively managed datasets to yield specific results useful for building

information products. These include data integration, summary ('aggregation'), statistical analysis (including spatial analysis) and other interpretative techniques such as modelling and forecasting.

- Production processes combine the results of analysis with other sources of information, such as the history and context of the issue concerned, and supporting details like acknowledgements and method of follow-up. Production also involves packaging and communicating information products which may require specific processes of its own, such as publishing and marketing (see Chapter 6).

The first step towards determining processing needs is to identify and describe the current processes and data flow. Formal 'process modelling' tools are available (see Section 3.3.5) to illustrate the flow of data and information between processes and to describe the processing which takes place at each step. Process modelling is frequently used by management consultants during quality improvement and re-engineering exercises. The objective is to analyse current business processes and suggest alternatives which enable the organisation to meet its output needs more effectively.

Process modelling may be applied at all levels. Thus whole agencies or departments may treated as processes in a high-level process model, and the resulting flow as evidence of partnership or linkage. High level process models are sometimes referred to as institutional linkages diagrams. They are a useful means of determining the co-ordination needs of multi-agency information systems.

Assuming that the objectives of the information system have been set (by an initial workshop or steering committee), it should be possible to study the current process model and decide what functions ('capacities') are missing. The capabilities of the agency (or agencies) concerned can then be examined and potential solutions proposed. One of three outcomes is likely:

- Current processes are adequate. Priorities must be set and resources allocated.
- Some processes are weak. Capacity enhancement is required, leading to the restructuring of weak processes (eg concentration of resources), provision of training or equipment, recruitment of new staff, or application of quality assurance procedures.
- Many processes are weak or poorly co-ordinated. Major capacity building is required to renew the agencies/processes concerned. Some processes may be replaced, enhanced or discarded if the opportunity to 're-engineer' the overall process is taken. Guidance may be required from international agencies and consulting companies.

The assessment of processing needs should lead to the following outputs:

- annotated process model (formal data flow diagram)
- institutional linkage diagram (as above but treating agencies as processes)
- description of the data management capacities of partner agencies
- description of the analysis techniques and related tools (eg software) employed by the above
- description of the desired outputs, including services and information products, of the information system
- assessment of the strengths and weaknesses of current processes, including suggestions for alternative process models and capacity building requirements.

3.3.5 Tools and Methods

There are useful tools and methods for determining and documenting user needs. Any particular assessment may require only a subset of these, the most appropriate methods depending on the depth of the study, the nature of the scientific endeavours, the organisational culture, and previous experience of the participants.

- Questionnaires

  Questionnaires are a highly structured method of data collection in which respondents are requested to 'fill in the blanks' on a form. This can be a valuable data collection tool in itself, or as a guide to facilitate data gathering, eg in interviews. A properly designed questionnaire promotes the *systematic* collection, cataloguing and evaluation of data. This eases the summarisation of general basic facts and trends. Data collection by this method is inexpensive and efficient.

  Questionnaires are best for collecting specific information or opinions on narrow options. The principal value is as a preliminary screening method to help determine which institutions or functions should be studied in more depth. As well, questionnaires can be helpful as a checklist or aide-memoire for conducting structured interviews.

  Questionnaires have limitations for open-ended or general assessment of user requirements and past experience has shown very low response rates are obtained from 'blind' distribution - that is, mailings without advance warning or explanatory material. Response rates can be improved by including a supporting brochure providing a summary explanation of the purpose of the study and questionnaire, together with a sample questionnaire completed as an illustration. However, even with this assistance, respondents may have difficulty answering some of the questions, may leave some fields blank, misinterpret questions, or bias answers based on incorrect assumptions.

- Structured Interviews

  The structured interview uses an independent person to obtain views through direct questioning and discussion. The interview is 'structured' in the sense that there are particular topics and/or questions which are asked in all cases, and standard explanatory information is provided in advance.

  Interviews may be conducted *individually* or as a *group*. Individual interviews can be conducted formally (questions are asked and responses recorded on tape or written down), or informally (questionnaire is used as a guide to discussing key topics).

  Information can either be recorded at the time or summarised following the interview. The interviewing approach should be sensitive to the cultural norms of the institution and individual concerned.

  Group interviews are useful where discussion and consultation are the preferred way to establish answers. A questionnaire or check list is used as a guide to solicit and record information. Information from the group is then summarised. In this approach it is useful to have one person to lead the discussion and another to record important information.

  Group interviews often benefit from a short presentation on the topic before opening up the discussion more fully.

- Working groups

  Working groups are small teams of individuals formed to address specific topics and return their results in a specified time frame. Working groups differ from Committees in having a time-limited mandate - no on-going role after the assigned task is completed. Working groups are usually composed of experts in particular fields rather than representatives of organisations. Working groups are a particularly useful way to refine information on a certain topic (eg a working group on indicators, or GIS) or to resolve serious problems or uncertainties.

- Workshops

  Workshops are similar to working groups in having the objective of addressing a particular topic. A workshop brings together relevant expertise for a short period (½ to 3 days) with the aim of producing agreement, better mutual understanding of issues, and a plan for future actions. Workshops often incorporate elements of training and, where a wide spectrum of institutions are involved, facilitate sharing of knowledge and expertise. Workshops always arrive at decisions by consensus.

- Brainstorming

  Brainstorming is a particular type of discussion technique in which the goal is to accumulate ideas on a subject in a short space of time. A facilitator is needed to initiate and steer the session, as well as create an atmosphere which stimulates creative thought. In a brainstorming session, all individuals are free to speak, and there is particular encouragement to put forward unusual and new approaches. All inputs are recorded. The ideas are then sorted and used where applicable in the context of the project. Brainstorming is most useful when defining the initial scope of a project, when a change in strategy is required, or simply for an infusion of new ideas and inspiration. For example, brainstorming may be useful in trying to identify key datasets in an institution, or new forms of information products to influence decision making.

- Data Modelling

  Data models illustrate the relationships between data entities, which may be defined as items of interest whose attributes (properties) are being recorded (see Section 4.5). The technique was first described by Peter Chen (1976). For example, an entity representing 'institutions' might be described by the following attributes: name, address, date established, mission, and annual turnover. The relationships between entities are depicted in 'entity-relationship' (E-R) diagrams, which use formal, consistent conventions to indicate different kinds of relationship. For example, a one-to-many relationship exists between an institution and its staff; and a many-to-many relationship exists between staff and the projects on which they work (assuming more than one person works on each project).

  Data models can be subjective. Thus two individuals may produce distinct but equally valid models of the same data, based on different sets of objectives for their applications. The kinds of data model most useful for user needs assessment are relatively high level (ie generalised), with more detailed modelling left until later stages of information system development (see Section 4.5).

- Process Modelling

  Following the methodology developed by Yourdon (1979), process models (or data flow diagrams) can be used to illustrate the flow of data between processes in a business or information system operation. A consistent diagrammatic convention is often used, with lines between processes and datasets indicating the direction of data flow. For clarity, it is conventional that each diagram should contain only a limited number of processes (4-6) and themes. The process model is useful in providing an clear overview of the existing operations of an information system.

3.4 APPROACH 1: STRUCTURED DEVELOPMENT LIFE CYCLE

3.4.1 Overview

A well established class of methodologies uses the Structured Development Life Cycle approach, in which the development is carried out in a series of structured incremental phases. Although different variants of the life cycle are advocated in different locations, all share the following basic features:

- there are distinct phases moving from conceptual issues to operation
- specific defined products result from each phase
- the phases are carried out in sequence, building on the products established in previous phases
- a decision as to whether to proceed is taken after the completion of each phase
- looping may be required to revise or refine products from the previous, but not earlier phases.

Figure 6 shows an example structure for the life cycle. Diagrams such as these have led to the 'waterfall' label being applied to this methodology.

The overall aim of system development is to create working databases in the agencies which are partners to the information system. Ideally, this is achieved using Development Teams drawn from the agencies concerned, rather than bringing in external consultants. However, the information system hub and, in particular, its Steering Committee may be required to facilitate system developments in other ways (eg training).

> *The structured development life cycle approach follows a series of logical steps from project initiation to operation.*

3.4.2 System Design

3.4.2.1 Purpose

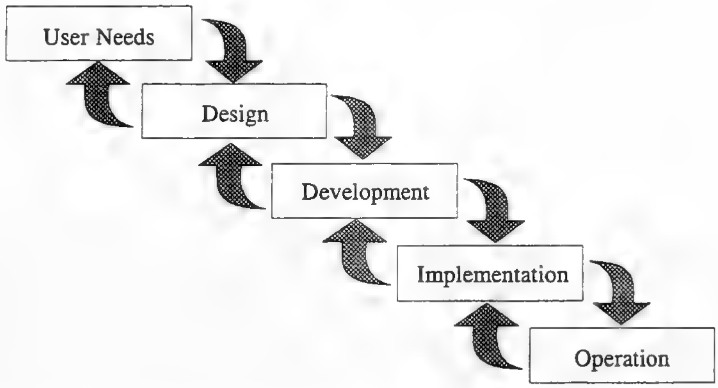
In the system design phase, the functional specification prepared in the user needs assessment is translated into a logical, and then physical design (see Section 4.4). This should result in a design based on the datasets and processes outlined in the functional specification which, once implemented, will deliver the outputs which users desire. The relationship between the different components of the information system (eg databases in different agencies) are made explicit in the design phase, and appropriate data exchange procedures are suggested.

Decisions are made on the overall system architecture during this phase, including the type of hardware and software to be used (see Section 4.6). The organisational environment has a large impact on how this is handled. The system may be implemented on existing hardware and software; but if no suitable equipment exists procurement may have to be initiated. The architecture of the system should, in most cases, enhance rather than replace existing mechanisms for data exchange amongst between different groups of users.

3.4.2.2 Activities

The Development Team puts together the system design in terms of the required data storage, access, and processing capabilities, and these are verified by selected users to ensure that they concur with user needs. Verification can be achieved by means of informal discussions, interviews and workshops (see Section 3.3.5), or by means

Figure 6: Structured Development Life Cycle



of prototyping techniques (see Section 3.5). Specific designs for sub-components, such as database applications, may also be undertaken at this stage (see Chapter 4).

The design phase provides an opportunity to begin training users in the system development strategy. If hardware and software are to be installed, effort is also needed to verify functionality against vendor claims, and to develop tight specifications for additional equipment and technical support.

3.4.2.3 Products

The product of this phase is a design specification which defines and prioritises the development tasks to be undertaken in the next phase, including details of any equipment required. Estimation of costs can be rigorous here, since these can be calculated by totalling the proposed development time and resources required (procurement costs can be confirmed by vendors).

The design specification should provide sufficient cost-benefit analysis to enable project managers to decide whether or not to request design modifications in order to satisfy time-scale or budgetary commitments.

3.4.3 Development

3.4.3.1 Purpose

In accordance with the design specification, database structures are established physically (see Section 4.7) and populated with test data to verify operation (see Section 4.8.1).

3.4.3.2 Activities

The major involvement is with the developers who are coding, testing and documenting the information system. However, user involvement should be maintained through demonstrations of functionality as they are developed. Continuing user involvement serves a number of purposes:

- assists with verifying, testing and debugging the system
- ensures that the system correctly addresses user needs (ie reflects the content of the functional specification)
- prepares users for delivery of the system in the next phase.

3.4.3.3 Products

The chief product of this phase is a functioning system which conforms to the design specification; the decision to proceed depends on this having been achieved. Assuming all is well, an implementation plan should be prepared for the following phase.

3.4.4 Implementation

3.4.4.1 Purpose

The purpose of the implementation phase is two-fold:

- to check the functionality of the system against user needs as laid out in the functional specification
- to establish and document effective operating procedures, including appropriate user manuals, data security policies, and data exchange guidelines for the system (see Section 5.4.2)
- to ensure that staff are familiar with these procedures by providing appropriate training.

The implementation plan produced in the development phase should guide how this is achieved. For instance, techniques for exercising the full range of system capabilities and administrative duties.

Functionality may be incorrect or missing, in which case details should be recorded for correction, and the affected parts of the system should be re-tested at a later stage.

The Development Team is often expected to absorb and implement a series of modifications during system testing. This should not be taken as an opportunity for users to demand fundamental changes in system characteristics, merely to check that their original needs are satisfied.

3.4.4.2 Activities

Both users and developers are involved heavily in this phase. The former organise and carry out system testing, and the developers correct, modify and fine-tune system performance. The results of this process should be recorded in the form of operating manuals, policies and guidelines.

3.4.4.3 Products

A functioning information system ready for operation, including the appropriate documentation, operating procedures and training provision.

3.4.5 Operation

3.4.5.1 Purpose

The operational phase is where the system should remain for its lifetime, becoming a regular feature of the agency or groups of agencies for which it was built.

During operation, users may detect errors in the system or conceive of improvements which could be made. It is important that a mechanism be put in place to accommodate feedback from users of this kind, in order to constantly improve system performance. One solution is the retention of a small technical support team (possibly some of the same individuals responsible for system development) who can respond to user problems and make changes 'on the fly' or during periods when the system is not actively in use.

The undertaking of major revisions or the correction of serious operational problems is best handled by the user community as a whole, via such mechanisms as a user support group or other forum.

3.4.5.2 Activities

Users review the performance of the system, including the documentation and suggested operating procedures, taking care to establish mechanisms for technical support.

3.4.5.3 Products

The outputs of this phase are those which are derived directly from using the system - ie the benefits of improved information management which were originally sought when the project was initiated.

3.5 APPROACH 2: PROTOTYPING

3.5.1 Overview

The structured development life cycle methodology described above has some disadvantages. The methodology requires a great deal of interaction with users in the early phases to define system requirements, followed by a (potentially long) period where the developers implement the specification. After this, the users once again become involved to test the final product. However, gaps in participation at any stage of system development can erode confidence in the Development Team.

Furthermore, user needs tend to evolve throughout the development period, making it essential to maintain dialogue on a regular basis.

With many industrial and administrative information systems it is relatively easy to specify the data requirements and the processes which are required to create the desired information. However, with biodiversity information systems (and many other scientific applications) the 'process' part of the specification is more difficult.

For example, it may be troublesome determining what types of analyses should be applied to the data, and how to summarise information in ways that are suitable to policy and decision-makers. This increases the risk that decisions made during the user needs assessment may need major revision.

These concerns have led to alternative, more interactive approaches to information system development which applies the concept of 'prototyping'. The principles of this approach are:

- to create a common ground between users and developers
- to have all parties understand the complexity of the processes being automated
- to build small versions of the system quickly (and inexpensively) so that user needs can be discussed in the light of a real example
- to allow changes to be incorporated easily during the development process
- to provide continuous interaction between users and developers throughout the development process.

The principal advantages are that the developers can quickly verify that their interpretation of user needs is correct, allowing problems to be identified and corrected early in the process.

Figure 7: The Throw-away Prototype

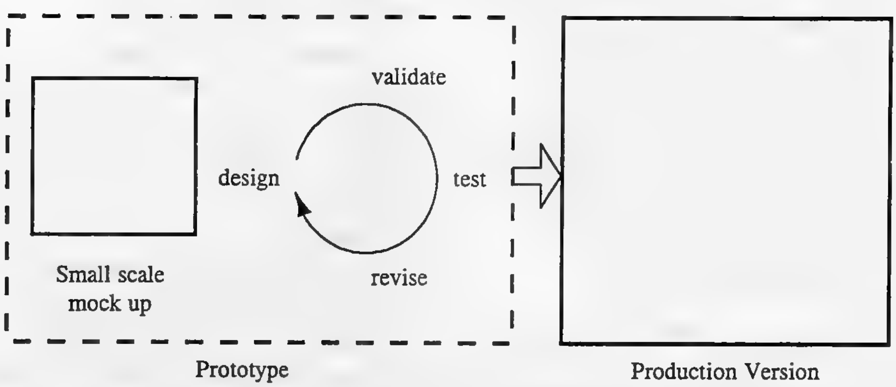
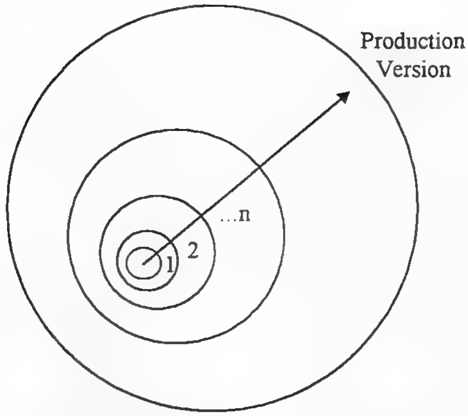


Figure 8: The Evolutionary Prototype



Prototyping methodologies develop 'mock-up' or partial systems within a short space of time, allowing potential users to provide feedback before proceeding.

Within this general framework, prototyping methods can be categorised into two types as described below.

3.5.2 The 'Throw-away' Prototype

With this approach a simple mock-up or demonstration of the system or one of its parts is built, demonstrating to users how it would perform in practice (for example, how the data entry screens would look, or how reports would be formatted).

The demonstrations do not necessarily use real data; nor are real analyses usually tackled at this stage. The prototype is rather like an artist's sketch of a new building (see Figure 7): it can be modified, perhaps several times, until users are completely satisfied, following which it is discarded and a real system (production version) is built.

3.5.3 The Evolutionary Prototype

The evolutionary prototype starts building a small part of the overall system (eg one process) all the way from design to implementation. Feedback from users is then incorporated into the design piece by piece, increasing the core capabilities of the prototype until it evolves into the production system. The result of the evolutionary approach is a system which can be adapted easily to future changes (see Figure 8).

3.5.4 Summary of Methodologies

The features of structured and prototyping approaches may be combined for maximum effectiveness. For instance, prototyping may be added as an additional phase in the structured life cycle, or applied during the design phase of the structured life cycle (see Figure 9). With combined approaches, adaptation to change is integral to the development methodology.

In practice, the traditional 'waterfall' approach works best for complex projects which are precisely defined in advance (ie high certainty of user requirements) and tightly controlled during development.

Conversely, prototyping works best with simpler, less easily defined projects, which may evolve as user needs are refined. A combination is recommended where the project is both complex and uncertain.

The choice of methodology and related tools is usually made by the team responsible for building or upgrading the system. Nevertheless, all users should be aware of the options in order to participate effectively in the project.

3.6 EXAMPLES

Two biodiversity information systems are profiled below. Further information on these, plus a range of other biodiversity application software, is provided in UNEP/WCMC (1995).

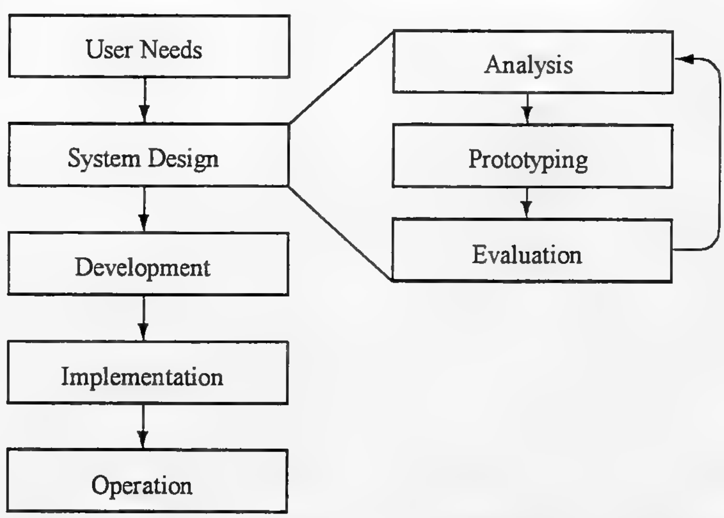
- BG-BASE

  An illustrative example of a computerised biodiversity information system is BG-BASE (see Figure 12), which was implemented following a request from IUCN to create a microcomputer-based application for botanical gardens, both large and small, based on the International Transfer Format (ITF) for plant data (see UNEP/WCMC 1995). A full account of the implementation process is given in Walter (1989), an excerpt of which is included below:

  "From the beginning the design of BG-BASE has been a group effort; it has now involved more than 100 people from over 35 institutions.... For approximately two years, a group of five to eight of us (specialists) met over lunch nearly every week to plan and to discuss the design, and eventually to test and criticise the implementation. Ideas for new data fields, new files, and new reports were presented regularly for general discussion, resulting in some fairly heated debates. The heart of the system was always understood to be based on the International Transfer Format, but since this format specified only 36 fields, we had a great deal of fleshing out to do. As it currently stands, BG-BASE comprises 564 fields spread over 12 major files. In addition to these major files, there are another ten index files that allow the user to look up information in a wide variety of ways"

The heart of the system was based, as requested by IUCN, on the International Transfer Format (ITF) for Botanic

Figure 9: Prototyping in design phase



Gardens, a protocol created for exchanging information (see UNEP/WCMC 1995). The value of using the ITF and the need to keep the application generic have become evident over time. BG-BASE has now been adopted by over 50 institutions world-wide to manage living collections, conservation information, herbarium specimens, and as a teaching tool. These institutions comprise botanic gardens, arboreta, horticultural societies, museums, universities and conservation monitoring centres.

The use of BG-BASE to manage plant conservation data at WCMC illustrates the importance of a flexible design. Although originally designed as a specimen-based system managing botanic gardens' living collections, BG-BASE has proved suitable for use in other contexts.

- Biodiversity Data Bank

Biodiversity Data Bank (BDB) was established at the Institute of Environment and Natural Resources, Makerere University, in early 1993, although the task of collating Uganda's biodiversity data began long before this using manual techniques (a full account is given in MUIENR/WCMC 1995).

The specification of BDB was conceived by a small Development Team with extensive knowledge of the information requirements of the biodiversity sector in Uganda. Many key organisations were consulted, including the Botany and Zoology Departments at Makerere University, the University Herbarium and Zoology Museum, Uganda Wildlife Authority, Forest Department, and several NGOs, such as IUCN and WWF.

The scope of the system is such that it can handle a wide variety of biodiversity data. This was considered important by users who requested a single system to manage their data, rather than a series of separate databases. The major data holdings include taxonomic names, species distribution records, protected area profiles, details of administrative units, a gazetteer, bibliography, and directory of contacts.

BDB was originally conceived as a means of organising the large amount of data relating to Ugandan biodiversity located inside and outside of the country. From the outset an aim of the system was species mapping, and thus facilities were built into the system to download species distribution data in a form suitable for desktop mapping programs.

However, due to a requirement to provide information on the country's protected areas system, pre-defined reports were also developed to list species, and in some cases estimate diversity, within protected areas. More sophisticated analyses were also developed to predict species distribution on the basis of observed habitat suitability.

4

Database Development

4.1 INTRODUCTION

Effective data management is central to the success of a distributed information system. The goal of this chapter is to promote techniques which inherently facilitate integration and exchange of data - thereby widening the range of applications for which the data can be used and simplifying the process of information production (see Chapter 6).

Custodian agencies should follow consistent, long-term methodologies for data collection and management, in accordance with the following principles:

- data should be collected and managed in their *primary* form, not classified, aggregated, or otherwise interpreted, allowing them to be used for multiple purposes
- data should be collected and managed following accepted *standards* (conventions) to reduce transaction costs and expedite interpretation by others
- databases should be developed and implemented using *generic* methodologies which facilitate adaptation to future needs
- databases should be implemented using *widely available* computer hardware and software to expedite access by others.

In addition to these core principles (which are elaborated later), a number of quality management principles should also be noted:

- datasets should be fully *documented* to facilitate use by others (see Section 5.3)
- *procedures* for operational and data security should be established (see Section 5.4)
- datasets should be maintained and used by *groups*, not individuals to increase operational security (see Section 5.5).

Data should be managed using standard, sustainable methodologies, which widen the range of applications of the data.

4.2 PRIMARY DATA

Environmental data record objects and phenomena in the physical environment. Some of these recordings are factual, for example the geo-reference of the location where an recording was made, the date of the recording, the dimensions of a tree, the weight of a log, the mean annual precipitation at a site, or the water retention capability of a soil profile. These are all *primary data* based on facts which can be measured against a stable, widely accepted standard (Busby 1994).

Secondary, or derived data are those developed from primary data by a process of interpretation or classification, either at the time, or later. Examples include: species name, vegetation type, canopy extent, and climatic zone. Derived data should not be stored in a database unless the primary data from which they were derived are also available. Why is this? Because, as concepts and paradigms shift, derived data are degraded in value and ultimately become useless. For example, if the only representation of a species distribution is an outline drawn on a map, this information becomes redundant if the species is split or otherwise disaggregated following a taxonomic revision. The correct approach would be to store the co-ordinates of the species observations (and supplementary identification notes) to enable new outlines to be derived.

The principle of storing primary data needs to be applied intelligently. No one, for example, would refuse to store the names of species or vegetation types, even though they are susceptible to change. The process of deciding which data to store is therefore one of risk assessment. Given the high costs of collecting data, the benefits of using a particular dataset should be balanced against the risk that it will become obsolete. As a rule, we should not be obliged to use data which are known to be deficient but which are too costly to replace or enhance.

The costs of data collection are particularly high in the case of large, national-level datasets, and strict priorities are therefore required for dataset production and maintenance. In general, it is wiser to develop nationally consistent datasets at low resolution, and progressively fine-tune, than to piece together more accurate, but inconsistent, local-scale datasets. This does not imply that local-scale data have no role in national information systems, only that a priority-setting framework should first be established to regulate their contribution (see Section 2.5).

4.3 DATA STANDARDS

Standards are the means by which people communicate information and are thus vital in any information system. Standards embrace the selection of attributes representing environmental phenomena, the nature and allowable values of those attributes, and how they can be used to greatest effect by stakeholders (Busby 1994). The purpose of standards is to lower the *transaction costs* of using data. Thus priorities for establishing standards should take into account the expected uses of the data, for instance in creating collaborative information products.

The development of standards requires a real commitment of resources, largely intellectual in nature. They cannot be overlooked, taken for granted, or left to a specialists who are not actively participating in the information systems project. They require concrete and determined attention by management; developing standards will not be easy.

Recognising that progress towards formally accepted national (and international) standards can be very slow, national information system projects will inevitably develop their own, interim, standards. In such cases it is vital to build on previous experiences, perhaps at the international or regional level, which may be available via international organisations and networks.

Interim standards are commonplace across many of the major themes, often having arisen to suit particular data collection and management objectives. Such *de facto* standards are propagated and adapted in local database implementations. The development of a multi-agency information system provides a good opportunity to reconcile and revise existing standards, taking into account a wide range of stakeholder's needs.

4.4 DATABASE DEVELOPMENT

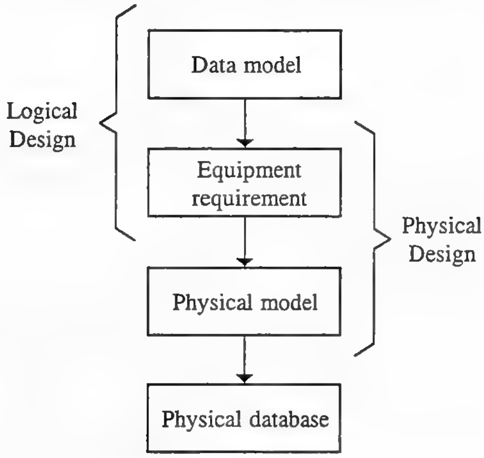
Database development involves designing and building the structures necessary to manage one or more related datasets. Generic methods are available to develop databases, and the ideas presented in following sections attempt to simplify and summarise these. The terminology for the following processes follows Daniels and Tate (1984).

A user needs assessment (see Section 3.3) is assumed to have taken place before database development is attempted. The assessment, which is written up in the form of a functional specification, is intended to provide all the details necessary to design the database in accordance with user's needs.

Database design is partitioned into two phases: the logical design phase, which is independent of the equipment used for implementation; and the physical design phase, which determines how the logical design will work using the equipment selected.

Linking these two phases is the analysis of the equipment required to implement the database, which in most cases involves the

Figure 10: Database development



selection of appropriate hardware and software. Figure 10 illustrates how these processes give rise to the final, physical database.

4.5 LOGICAL DESIGN

Logical database design involves identifying key datasets and studying how these need to be accessed and analysed to achieve the desired objectives. The logical design is independent of both hardware and software, and does not assume any particular method of physical data organisation (in practice the hardware and software platforms available - perhaps constrained by budgetary limitations - may affect the final logical design).

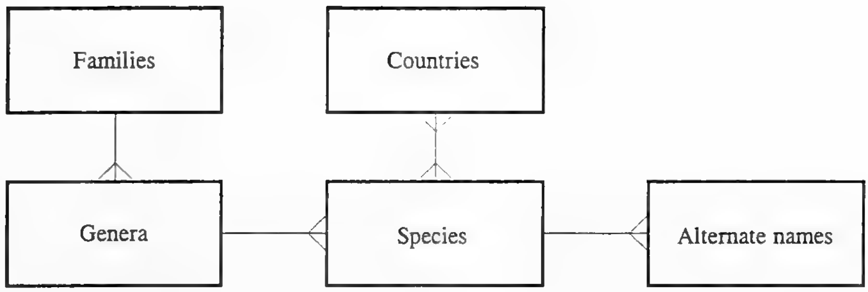
The advantages of producing a logical design are:

- it provides a stable base from which to set standards and co-ordinate the development of the database
- it provides a conceptual model which is completely free of implementation considerations, and which can be used as a point of reference when adding to or modifying the functionality of the database, or changing the equipment on which it is based
- it provides a specification which can be used in the evaluation of alternative data management software
- it provides a base line from which an optimum physical data organisation can be produced.

It is important for users to achieve a common understanding of the datasets managed by an agency - ie those required to meet its 'mission-critical' needs as identified in the user needs assessment. The process of the structure and inter-relationships between a group of datasets is referred to as data modelling, and various language and diagramming aids exist to standardise this. The process of data modelling is facilitated by dialogue with domain experts who are familiar with the dependencies and inter-relationships between the major themes.

The first step in the development of a data model is to study the functional specification resulting from the user needs assessment. Consideration of this document, together with discussions with both users and experts, permits determination of the basic 'items of

Figure 11: E-R model



interest' and hence the initial entities of the data model.

The next step is to determine what relationships exist between the entities that have been identified. It is important at this stage to concentrate on the 'natural' relationships which exist, rather than just those which it is thought may be computerised.

Data models are often represented in a formal manner. The most popular representation is the entity-relationship (E-R model), first described by Peter Chen in 1976. This model provides a very clear diagrammatic representation of the top-level objects to be modelled in a domain. In the original paper, Chen set out the foundation of the model; it has since been extended and modified by Chen and many others. In addition, the E-R model has been made part of a number of Computer Aided Software Engineering (CASE) Tools (see UNEP/WCMC 1995). Today, there is no single E-R model, although most share the features outlined below.

- Entities

  Items of interest (concrete of abstract) whose attributes are being measured. Entities are represented as tables in a physical database.

- Attributes

  Properties of an entity which are measured to produce data (eg 'designation' is an attribute of the 'Protected Areas' entity). Attributes are represented as columns or 'fields' in database tables, such that all instances of a given entity are structured similarly.

- Relationships

  Descriptions of how two entities relate to one another (eg 'species' may be related to 'genera' by a 'belongs to' or 'many-to-one' relationship). Figure 11 illustrates this.

Note that alternative symbols may be used to construct entity-relationship diagrams. The notation adopted in this document follows that of Ashworth and Goodland (1990). Connecting lines between entities are single or forked depending on their relationship, forked lines indicating the 'many' side of a one-to-many or many-to-many relationship (see Kroenke 1992 or UNEP/WCMC 1995).

The advantages of producing a data model are:

- improved dialogue between users, and consequent development of data structures
- identification of redundant data
- improved capacity to identify data validation criteria

- a formal, possibly automated method for implementing the physical database.

Prepare Entity-Relationship diagrams to explore data relationships and record the data model.

4.6 EQUIPMENT REQUIREMENT

4.6.1 Overview

Following the data modelling phase, the next step is to study how the data will be used in practice. This involves analysing what kind of integration, analysis and communication processes will be applied to the data, with the intention of deciding what kind of data management equipment is required.

Before embarking on the potentially costly process of selecting computer hardware and software, it is worth deciding whether or not such equipment is actually justified. Some advantages of the latter include the ability to:

- enforce consistency and structure in data storage, which contributes to data quality
- automate validation during data entry
- analyse large volumes of data
- produce multiple and varied reports from the same data.

Developers considering whether to invest in data management software should ask the following questions:

- do the data contain relationships too complex for the capabilities of a manual filing system or word processor?
- will the quantity of data be too much for manual methods or word processing to efficiently handle?
- will it be necessary to integrate data from several sources into a combined output?
- will there be a need for the data to be shared amongst more than one user in a single institution, or with other institutions?
- do the data require extensive searching, sorting, or updating?
- will frequent reporting of the data be required?

If the answer to some of these questions is yes, then the use of specialist data management software should be considered. If the answer is yes to many questions then such software is certainly required.

Evaluate whether a special-purpose computer system is required before proceeding.

4.6.2 The Selection Process

Assuming that computer hardware and software are required, the following questions about the database should be answered in order to specify the need:

- How big is the database? How many individual entities will be included? How many cases (instances) of each entity are there?
- Are any special data types needed, such as spatial data, large volumes of text, images, sounds, or video? Will document storing and searching be necessary?
- How many people need access to the database? Will they be sharing a single computer or using a network? Are they all in the same institution or physical location?
- What are the long-term plans for the database? Will the scope or the number of users grow?
- How much computer experience does the implementing agency have (eg for technical support and maintenance)? How much time is there to learn new software?
- How much money is available to spend on hardware and software?

4.6.3 Software

The most commonly used form of data management software on the market today is the relational database management system (RDBMS). These offer good flexibility and performance at modest cost, although they do not deal easily with large-scale textual sources (see Section 6.3.4). Many evaluation criteria can be used to select a suitable data management package, some key examples of which are given below:

- is it powerful enough to manage the expected volume of data?
- will it meet user expectations in terms of look and feel?
- does it contain good facilities for applications development? (the amount of money spent on applications development usually exceeds the initial costs of the software, so short development periods can result in significant savings)
- is it a popular product which will continue to be supported and enhanced? (it can be beneficial to forsake the latest technology for the stability and support of a well established product).

The above criteria should be evaluated against the requirements of the physical database design. However, counting the number of check marks in each case is a poor way to compare products, since key features like speed and reliability overshadow lesser capabilities. Ideally, the software is tested under realistic local conditions. Published software 'benchmarks' are often optimistic and may not reflect the demands of the destined database. Many important software characteristics are subjective. These include ease of use, consistency of the user interface, and expressiveness of a programming language. Selecting a software package purely from a list of features is unlikely to be satisfactory; nothing can substitute for examining a live installation.

Reputable computer magazines often contain advertisements and wide-ranging reviews of software packages, although these too can be biased (software reviewers sometimes have connections with vendors whose products are under review). If you rely on published reviews, temper the prejudices of any one reviewer by using several sources.

Computer bulletin boards are another source of outside expert advice. The vendors of very popular software packages usually maintain bulletin boards which may be accessed via services such as Internet newsgroups and CompuServe Forums. Bulletin boards not only store objective assessments of software, but can also provide solutions to technical problems via a network of remotely connected users. Knowledge can often be gained simply by observing the debates and comments of other users.

When selecting a software package consider the criteria that are of most importance to the project; prioritise these and then assess how well different products perform.

4.6.4 Hardware

Depending on the capability of existing hardware to support the desired design, and the availability of resources to acquire further equipment, new computer hardware may be commissioned to implement the design. Common architectures for this include:

- stand-alone computers
- locally networked computers with database software residing on a file server machine (LAN)
- client-server architecture
- a fully distributed database consisting of a series of remotely networked computers communicating via permanent or dial-up communication lines (WAN)

The third option, client-server, is becoming an increasingly popular solution to the data processing needs of medium to large-sized organisations. This architecture is a hybrid of the stand-alone and the traditional network options. It integrates the best characteristics of personal computers (friendly software and quick response) with the best traits of powerful centralised servers (high storage capacity, data exchange, strong security). The client-server architecture divides tasks between the user's computer running 'client' software, and a central computer running 'server' software. Typically, critical datasets are stored on the server where they are managed very securely and can be processed at great speed. The client software (running on the user's computer) sends requests to the server software when data processing is required. The processing then takes place on the server and the results are sent back to the client. Many clients can communicate with the server at once, allowing flexible, yet highly secure, data processing.

Key issues to bear in mind when selecting a suitable platform are:

- Scaleability

  As the number of users, records, or features, grow, an application that once performed perfectly well on a low-cost architecture can drop off in performance quickly. Typically, stand-alone or small network computer architectures are most likely to suffer from this problem, which explains the rise of more sophisticated architectures such as client-server.

- Connectivity

  To enable rapid exchange of data between individuals and agencies, electronic connectivity is very desirable. This could take the form of a group of locally networked computers sharing a common storage area, or more sophisticated dial-up communication lines to external services such as the Internet and private networks. The capacity to connect computers together is becoming increasingly recognised as the key to rapid dispersal and exchange of data.

- Compatibility

  The issue of hardware and software compatibility is now diminishing in importance as manufacturers evolve a range of 'standard' specifications for their products. However, the so-called standards are still too varied and numerous to discount the problem entirely. As far as hardware is concerned, the major decision on compatibility is whether to adopt IBM-PC compatible computers, Macintosh computers, or (usually) larger workstations running the UNIX operating system. Within this broad classification, issues such as operating system choice, emulation software availability, network operating system (eg Novell, Vines, Lantastic), connectivity protocols between databases (eg ODBC) tend to dominate. At all stages, the best solution is to adopt technology which has been proven to be reliable and useful in circumstances similar to those anticipated, working on the principle that in such cases, compatibility issues are unlikely to cause serious disruption.

When selecting computer hardware, attention should be paid to its scaleability, connectivity and compatibility with existing equipment.

4.7 PHYSICAL DESIGN

Physical database design involves adapting the logical design to the requirements of the equipment used for implementation.

Transformation of the logical design into the physical design is usually straightforward: entities in the data model become a *tables* in the physical model, and attributes become table *fields*. The way in which relationships between the entities are dealt with depends on which data management software is used (see Section 4.6.3). If the chosen package does not support some types of relationship, then this has to be resolved by altering the logical design.

Each field in the database should be documented in terms of its purpose, data type, size, and order in its corresponding table. When pooled across all the tables of the database, these definitions are known as the *data dictionary* of the database, and provide a complete description of its structure, format, and use.

The business world is highly heterogeneous and a database for one company is unlikely to use the same data dictionary as that of another. In contrast, it is likely that countries and organisations managing biodiversity data may be recording and tracking many of the same parameters. Thus in the interests of data exchange and co-operation with external partners, notice should be taken of existing standards and common practices (see Section 4.3).

There are currently several international projects to assemble environmental thesauri (see UNEP/WCMC 1995). These are being developed in multi-lingual versions (primarily European languages at this stage). The most mature of these thesauri is the *INFOTERRA Thesaurus of Environmental Terms* (UNEP 1990), which currently contains around 1,600 terms. This number is not sufficient to cover many local terms, and must therefore be augmented in such situations.

During the transformation of large databases from logical to physical design, CASE tools (Computer Aided Software Engineering) can prove useful. These allow E-R diagrams to be drawn up, and used to validate and maintain the logical database design. Some CASE tools are also able to output the E-R diagrams directly into a Data Definition Language (DDL) that prescribes the physical database design.

Compile a data dictionary for the database using standard terminology and thesauri where possible.

4.8 IMPLEMENTATION ISSUES

4.8.1 Data Entry

Following completion of the physical design, the latter is transferred to the selected hardware and software (see Section 4.6.2) by creating appropriate database tables. The next step is to *populate* these tables with the required data.

Ideally, all the necessary data have been computerised previously and are available in electronic format for importation into the database. However, data are frequently in the wrong format or available only in hard copy form. In such cases they must be converted into an appropriate form for importation or entered manually into the database via the keyboard or other input tool (eg a scanner or digitising tablet in the case of maps).

Custom programs can be designed to regulate and validate data entry in many database and spreadsheet packages. This idea can be extended to automate other processes such as querying and reporting data, and 'downloading' data for exchange. A database which is accompanied by automated data entry or other procedures is often referred to as a database 'application'.

Where data are entered via the keyboard, validation checks should begin with rigorous examination of the raw, normally hard-copy, data sources. This can be a labour-intensive and tedious task, but is very important for maintenance of data quality (see Section 5.3). Where data are not entered directly, but are imported from another electronic source, validation checks should be performed on all the imported data. As an illustration of how errors can be introduced into a database by manual typing, suppose that a data entry screen has 10 fields, and that each field takes on average 6 characters to fill. If the success rate of the typist is 99%, then the chance of the whole screen being entered correctly is $(0.99)^{(10x6)}$, which, surprisingly, is only 55%.

An example of the types of validation check applied to species distribution records prior to entering a database is presented below (Richardson 1994):

- records are checked to see that all required data fields are present
- scientific names are checked for validity
- grid references of terrestrial species are checked for being over land, not water
- the presence of a species in a certain location is tested against a prediction based on bioclimatic factors, and outliers selected out for further validation.

For large applications, it is a good idea to write special-purpose validation routines, or take advantage of automated procedures offered by most data management software. Such routines perform 'reasonableness' checks on field values, such as ranges for numeric fields, or string-length for character fields. It may also be possible to enforce consistency checks such as capitalisation and hyphenation. Finally, many packages permit the user to select values from a set of predefined choices. This eliminates the possibility of typographic errors, and can speed up data entry considerably.

Data validation procedures should be established to reduce errors during database population.

4.8.2 Synonyms and Equivalent Terms

In a typical data management package, data are retrieved by means of structured requests or 'queries'. Thus, if the user wants to find information on protected areas by providing the search string 'protected area', the search will fail to retrieve records marked 'park', or 'reserve' or 'sanctuary', despite the semantic similarity. The problem of synonyms and equivalent terms is particularly prevalent in the environmental domain due to its heterogeneous make-up.

This difficulty can be overcome by developing custom search routines using the facilities of the software, and offering them to the user as menu or push-button options. An on-line thesaurus can also assist the user by providing a series of alternative search terms. This can be done in a passive mode by suggesting the terms to the user on request, or in active mode where the thesaurus is automatically consulted during the search process to identify synonyms and semantic matches.

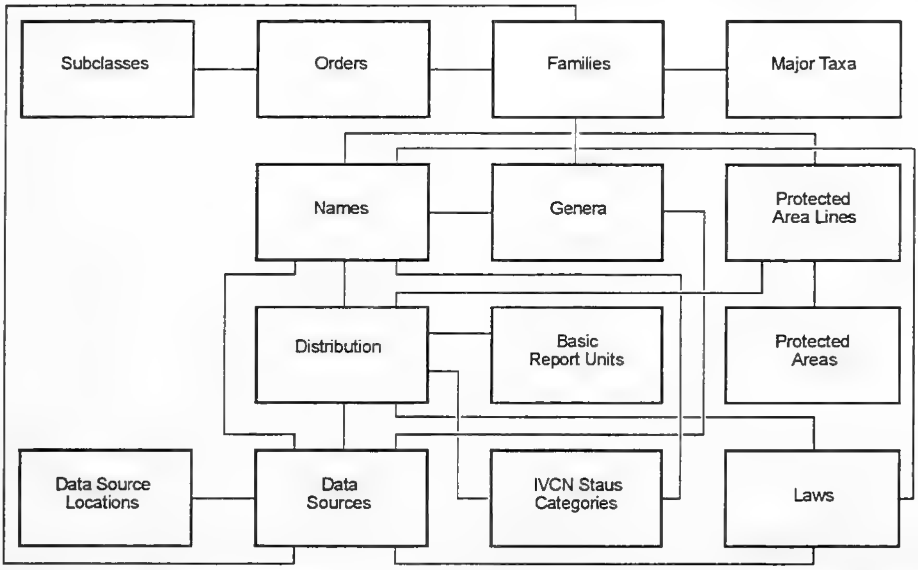
4.8.3 Hierarchical Data

Hierarchical structures are required to manage many forms of biodiversity data, including species names (order, family, genus, species), geographic relationships (region x is located in country y, in continent z), and other multi-level classification systems used for the description of land use, vegetation, and other ecological units.

In a recent study from Australia, Richardson (1994) highlighted the problems encountered when establishing a taxonomic database structure, and the need for these to be tackled during early stages of the system development process. The same kind of problem, which arises when an attempt is made to manage data which may not be formalised, complete, or even agreed, occurs similarly in the case of habitat or ecosystem classification categories.

Firstly, systems had to be designed to integrate differing standards between disciplines (eg botany and zoology) and between institutions. This is especially common at the generic level where different practices can result in the 'splitting' or 'lumping' of genera. Secondly, taxonomic standards change with time, as knowledge of the phylogenetic relationships between

Figure 12: E-R diagram showing the relationships between tables in BG-BASE



species improves. Thus data supplied by different sources may use differing names for the same species, and the database structure must be able to integrate these synonyms. This situation may also arise when it is discovered that taxa previously thought of as one species consist of two or more and, as a result, a part of the data for a species is included under the wrong name. Richardson suggested that taxonomic database structures should take into account the following:

- Formal Categories

  The family, genus, species, sub-species, other infra-specific categories, and corresponding authorities of the taxa (family name is included as the same name may be used for genera of plants and animals).

- Applied Categories

  Users may need to associate other names with the formal categories, such as synonyms and common names. Applied categories should be fully referenced in terms of authority, date and source.

For example, in the BG-BASE database used at WCMC to store data on threatened plants and plant collections, plant names are stored in a five-tier hierarchy comprising the Names, Genera, Families, Orders, and Subclasses tables (see Figure 12). Note that a sixth table containing synonyms (not shown) is linked to the Names table. The hierarchy described stores plant names with minimum storage overhead and, with properly structured reports, can be used to respond to queries such as 'list all distribution records of species belonging to the same family as Acer palmatum'.

5

Quality Management

5.1 INTRODUCTION

Quality management refers to the overall process which governs the quality of a product from beginning to end. In the case of information production, the process begins with data collection and ends with use in decision-making. Quality control checks and quality assurance methods may be applied through all stages of this process.

There are no absolute measures of product quality. What may be 'high quality' for regional planning may be poor or useless for local decision-making because of factors such as scale, detail, and error. Environmental data, particularly in biodiversity, are rarely free of errors or '100% accurate', as they may be drawn from subjective observations (eg deciding the boundary of a habitat), incomplete sampling procedures (eg inventory work), or indirect measurement (eg remote sensing).

Even if it were theoretically possible to manage complete and accurate environmental data, time and cost considerations would prevent this in practice. Thus, with rare exceptions, it must be assumed that *all* biodiversity datasets contain errors and uncertainty. In such circumstances 'quality' becomes a measure of 'fitness for use' - ie dependent on its proposed use. This is important to remember when data are requested for uses different from their original purpose.

Product quality can be improved by attention to all aspects of institutional and data quality management. In the long-term a product's quality is judged by its users, and thus serious attention should be given to user needs and user satisfaction - the so called 'end user' approach.

A product's quality is a measure of its 'fitness for use'; many aspects of institutional and data quality management affect product quality, including attention to user needs.

5.2 INSTITUTIONAL QUALITY

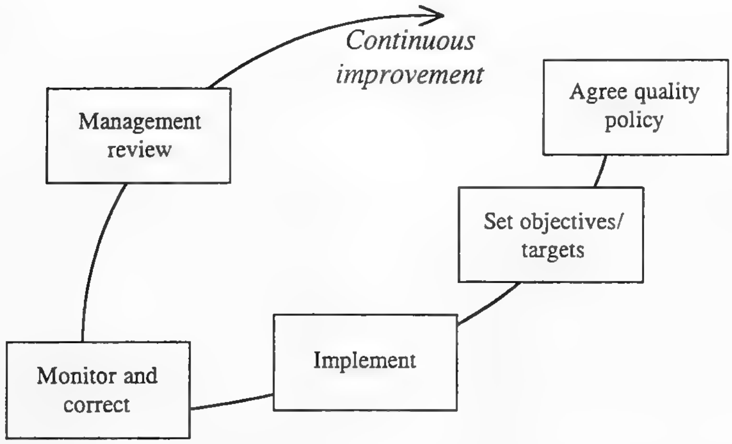
The establishment of institution-wide quality standards is exemplified in the series of quality management standards of the International Organization for Standardization (ISO), referred to as 'ISO-9000'. These standards are generic and process-oriented; that is they do not specify any specific levels of quality for products, but instead insist on a process of *continuous improvement*. This develops institutional performance, not necessarily in all areas simultaneously, in line with a well-defined quality policy (see Figure 13). Active participation is sought from staff to ensure that quality deficiencies, and quality management deficiencies, are diagnosed and treated.

Much simplified, ISO-9000 requires an institution to provide:

- a quality policy that everyone in the institution should understand
- a method of measuring the quality of information outputs that is applied consistently
- a method of determining external user satisfaction with information outputs that is applied consistently
- a feedback mechanism which ensures that internal and external measurements are actually used to ensure or improve the quality of the information service, as specified in the quality policy.

The overall emphasis is on the end user (client), and on quality considerations across all aspects of the operation (hence the term 'Total Quality Management' is often applied). The organisation is free to establish its own quality policy, specific quality measures and targets, measurement methods and feedback mechanisms which are appropriate to the needs and the nature of the issues being addressed.

Figure 13: Quality improvement loop



The ISO-9000 series of standards do not explicitly include environmental performance as a quality measure, despite the growing need for organisations to reduce negative impacts on the environment (the commercial, governmental and non-profit sectors are all responsible).

The continuous improvement approach is therefore being applied to, amongst other topics, environmental management systems and 'environmental auditing'. This will result in a 'greener', more sophisticated series of standards known as ISO-14000 (ISO Technical Committee 207) during 1996.

While quality management systems such as ISO-9000 deal with customer needs, environmental management systems address the needs of a broad range of interested parties and the evolving needs of society for environmental protection (British Standards Institute 1994).

Recognition that an organisation has complied with the processes and conditions advocated by ISO-9000 or ISO-14000 is achieved via a process known as certification. This is normally conducted by an independent third party (according to the ISO-10000 Standard for example) which audits and certifies that the organisation has established specific processes with regard to their products and services, and that they are being actively followed.

Some institutions in the biodiversity sector may wish to seek certification, but this can be a major and costly undertaking. It is suggested, therefore, that institutions implement a quality management process which follows the spirit of ISO-9000, seeking certification as and when this is feasible.

In the interests of environmental protection and cost reduction, it is also recommended that elements of the ISO-14000 series of standards on environmental management systems (EMS) are reviewed and implemented.

Agencies should implement a quality management process following the spirit of ISO-9000, seeking certification as and when this is feasible.

5.3 DATASET DOCUMENTATION

5.3.1 Overview

In the past, agencies rarely devoted much attention to data quality. This was because datasets were usually built for one specific project by people who well understood the nature of the data, including its deficiencies and caveats. At the end of the project the dataset was usually archived, filed, or neglected. Although regarded as desirable, dataset documentation has seldom been accorded a high priority because no one believed it would be of much real value.

Because datasets can be used for multiple purposes within an information system, comprehensive documentation of datasets is increasingly being recognised as an essential obligation of data custodianship and, in addition, a strategic corporate asset. Indeed, the preparation of dataset documentation should be planned thoroughly - including suitable allocation of resources.

The results of a documentation exercise, which might include an assessment of uncertainty or limitations in a dataset, its original source and intended purpose, are collectively known as 'metadata' or 'co-data' - ie data about data.

5.3.2 Metadata

A metadatabase record should contain the information needed to correctly interpret and use the data. Elements of this include:

✓ details of custodian - institution name, address, contact person

✓ data structure, format and media

✓ data collection method(s) - recording technique, equipment used

✓ access - available formats and media, cost, restrictions on use

✓ history of original sources (if secondary)

✓ data interpretation techniques applied

✓ data dictionary - definition of attributes, coding schemes and 'standards' employed

✓ intended use(s)

✓ data quality - quality assurance procedures applied, quantitative quality estimate, qualitative quality statement, including known limitations or deficiencies

✓ geo-referencing information (for spatial data) - projection, origin and offsets

The fundamental principle in metadata development is 'truth-in-labelling'; that is the dataset should be exactly as described and of a quality which is suitable for its stated (and implied) uses. Quality 'audits' of important datasets should be undertaken periodically, with particular attention to the completeness and accuracy of 'metadata'.

Datasets should be documented following the principle of 'truth in labelling'; the resulting metadata should be audited periodically.

5.3.3 Spatial Data Quality

Storage of spatial data imposes additional responsibilities on data quality managers. Three common questions to ask of spatial datasets are given below:

- are the data a faithful reproduction of the original source? (if digitised from paper or copied/transformed from an earlier digital source)
- to what extent is this an accurate representation of the spatial phenomena? (ie how does it match reality on the ground, with associated questions of resolution or effective 'scale')
- are the data internally consistent?

This last question seeks to ensure that data elements are consistent with both themselves and the stated topological and structural constraints. Basic tests for this are listed below:

✓ all polygons closed

- ✓ networks correctly joined
- ✓ left and right pointers correct
- ✓ one-to-one relationship of spatial objects to attributes
- ✓ edge matching of 'tiles' correct (both attribute and spatial)
- ✓ natural phenomena represented consistently, eg streams flow down hill, rivers are in valleys, identified peaks are at the top of hills, cultural features consistent with land cover
- ✓ data missing in some layers but not others.

Minimum standards should be defined for each of these quality areas (at least of the 'must be present' variety). These should be documented and made available to users of the service (similar to the *Quality Manual* of ISO-9000).

5.4 OPERATIONAL AND DATA SECURITY

5.4.1 Overview

A range of operational and data security procedures are required to guarantee data integrity on a day-to-day basis. In particular, data should be protected from accidental erasure which may occur due to:

- human errors in copying files, updating records, reorganising databases, and other operational procedures
- mechanical failure of disk drives and logical faults caused by power failures and fluctuations occurring during database transactions
- destructive effects of computer 'viruses'.

In general, threats to data security tend to be greatest where:

- the physical environment is hostile to computing equipment (eg extremes of temperature, high humidity or dust)
- electronic interference is strong (eg hospitals, industrial plants, locations near transmitters)
- power supplies are uneven or unpredictable
- informal (virus-prone) computer networks are the primary means of data exchange.

Operating procedures (protective measures) can be introduced to help combat the most common data security threats. Effective procedures include:

- regular (eg daily, weekly and monthly) backup of all critical data on removable electronic media (eg magnetic tape, optical disk)
- storage of backup media 'off-site' - ie away from the workplace in order to restore data after damage or theft of key equipment
- periodic 'test' restoration of backed-up data to ensure the procedure is straightforward and effective
- periodic 'test' recovery from simulated virus attack, hardware malfunction or other disaster
- regular virus checking with up-to-date software
- avoidance of unlicensed or 'borrowed' software, computer games, or other personal software
- power regulation via the use of uninterruptable power supplies (UPS), surge protectors, and radio interference filters.

5.4.2 Implementing Operating Procedures

Operating procedures, including those outlined above, should be documented in User Manuals, Data Security Policies and Data Exchange Guidelines, so that all users have the chance to review and understand them.

- User Manuals

  Typical procedures outlined in a User Manual might include methods for starting, using and exiting desktop applications; tips on data integration and analysis techniques; potential pitfalls; and case studies illustrating how to use the application for maximum effectiveness.

- Data Security Policies

  These might contain details of minimum backup requirements; power regulation requirements; procedures for avoiding computer viruses; and general regulations to ensure the integrity of the workplace is maintained. Specific plans might be included to recover from emergency situations such as virus attack, hardware malfunction, fire or theft.

- Data Exchange Guidelines

  Data exchange can be beneficial to both provider and recipient of the data. However, there is a need to establish operating procedures to make sure that:

  1. contributing data sources and intellectual property are properly acknowledged
  2. release of the data does not put biodiversity at risk.
  3. appropriate documentation is included (eg a summary, key, quality statement, guidelines on use)
  4. the transaction is financially sustainable for both provider and recipient (costs are recovered by provider, recipient can afford data)

The main obstacle to documenting and implementing operating procedures is normally shortage of trained staff and resources. Nevertheless, management should accord a high profile to data security, irrespective of the resources available, to encourage personal awareness and ownership of the problem. On occasion an entire institution or programme can be forced to close due to loss of critical data. This occurred in the South Pacific when a freak wave struck an office, eliminating its data. No copy of the data was maintained off-site.

A full discussion of operating procedures, which are a basic requirement of professional information management, is beyond the scope of this text. However, key references relating to data quality are provided in the UNEP/WCMC (1995).

Document and make widely available procedures for operational and data security, including user manuals, operating guidelines, and policies for backup, virus protection, and disaster recovery.

5.5 HUMAN RESOURCE ISSUES

5.5.1 Overview

Computer technology is marketed as a time-saving solution to a wide range of scientific, production and secretarial tasks. However, to achieve the desired improvements in efficiency and effectiveness, a high level of professional expertise is required. Three broad areas of expertise can be defined:

- technical support (including systems management)
- information production
- strategy development

Since the number of people possessing these skills is often low, recruitment at small or remote sites may be problematic; the individuals are in demand in large enterprises, particularly in the financial sector, which offer higher salaries, peer interaction, training opportunities and career advancement.

The challenge of recruitment and maintenance of qualified staff is a fundamental quality management issue - especially with regard to operational and data security. Indeed, the only way in which an organisation can build resilience to staff turnover is by making sure critical datasets are maintained and used by groups, not individuals. The way to achieve this is empowerment of staff by regular, effective training.

Every item of new technology acquired by an organisation creates an additional training need, which implies direct costs if staff are sent on training courses, or indirect costs if staff lose productive work time while acquainting themselves with the new technology.

5.5.2 Technical Support

Technical support staff have two major roles:

- to maintain and develop a secure and productive computing environment in which users can undertake tasks without worrying about technology
- to advise, guide, and train users in the use of different components of the computing environment

Depending on local conditions, experience shows that an ideal ratio of technical support staff to users lies in the region of 1:10 to 1:50. Alternative sources of support, such as telephone hotlines and support services, may assist in specific circumstances but should not be relied upon. However, the provision of manuals and computer books is strongly recommended as a mean of developing user awareness.

Typical technical support skills include:

- network operating systems (eg Novell Netware)
- general purpose packages (eg word processing, graphics)
- data management packages (eg DBMS, spreadsheets)
- information production tools (eg hypertext, multimedia)
- communications software (eg email, Internet, router, bridge software)
- scientific packages (eg GIS, data analysis, modelling).

Some of these skills may be provided by equipment suppliers in the form of technical representatives whose services form part of a contract; some may be provided under contract from specialist consulting firms (eg database design specialists); others may be available in-house.

Consider different approaches to obtaining specialised technical support, including contracting, sharing expertise with other institutions, as well as reliance on in-house staff.

5.5.3 Information Production

As we saw in Chapters 2-4, the development of information infrastructure presents a wide range of organisational and technical challenges. Key issues include system design and development, adherence to standards and quality assurance procedures, and production of information for different classes of audience.

The kinds of skills necessary for effective information production are:

- solid background in data management and information production concepts and technologies
- ability to analyse and present development options to cross-disciplinary teams
- flexibility to adapt development plans in accordance with user needs
- creative approach to product design issues such as content and layout.

Evidently, a core competency in information technology is required. However, the ability to work in multi-agency, cross-disciplinary teams is also vital, as is an appreciation of good design.

5.5.4 Strategy Development

In order to achieve lasting improvements in quality management, all institutions, particularly those actively managing data, should work towards strategic information management objectives. The development of strategies within an institution is normally undertaken at the senior management level. It should nevertheless be conducted with the full participation of more junior staff, especially those who are technically literate or have specific skills.

The kind of skills necessary for information strategy development are:

- clear vision of the mission and direction of the institution
- familiarity with the potentials, if not the details, of information technology
- realistic understanding of resource availability
- expertise at user needs assessment, both of internal staff and external clients.

Clearly, these skills are more management oriented than scientific or operational. However, a good information strategy will not necessarily maintain an institution at the 'cutting edge' of technology, but will enable it to apply technology effectively to improve the quality of its products.

> *Steadily improve institutional quality by developing and working towards an information strategy.*

5.5.5 Professional and Vocational Standards

Many graduates of universities and technical colleges in scientific fields acquire competence in information systems and feel comfortable with computers. However, they may never have held responsibility for designing or trouble-shooting information systems, and may be unable to function effectively in operational situations.

It may be a long time before major training institutions (such as universities and colleges) are able to provide graduates qualified for biodiversity data management. Until the subject has matured, specialist training may be necessary to maintain institutional quality standards.

The appropriate approach to human resource development will depend on such factors as stability and duration of tasks to be undertaken; local availability of skills; obligations of suppliers to provide support services; institutional staffing budgets; and partnerships with international organisations for training and related capacity building needs.

> *Develop cost-effective training strategies for different tasks, including technical support, information system design and product design, which take into account the practical options available.*

5.6 EXAMPLE

Insight into the application of quality management to biodiversity information management can be obtained by reviewing the procedures of experienced organisations.

An example is the Environmental Change Network (ECN) which has a long-term monitoring programme at a large number of sites in the UK. ECN structures its data using the Oracle RDBMS. Datasets are fully documented within this system, including a quality assessment and quality code. Detailed 'Measurement Protocols' are provided to data gatherers during monitoring operations, helping to ensure that data are collected consistently and that factors which influence measurement quality are recorded. Overall quality policies and objectives are being defined in the spirit of ISO-9000.

More information on this organisation and others involved with biodiversity information management may be found in UNEP/WCMC (1995).

6

Information Production

6.1 INTRODUCTION

Sound decision-making relies upon a good understanding of just a few key facts and issues at any point in time. The implication of this is that information should convey simple, succinct messages which influence decisions and achieve change.

Decision-making processes can be highly variable amongst different groups and cultures, and the method by which information reaches its audience can affect its impact significantly.

With issues competing for decision-making attention in an information-crowded world, timing can also makes a dramatic difference to the way information is received. Even the simplest, best presented information will have little impact if its message has run out of steam.

In essence, information is most effective when it is clear, timely, and delivered in recognised ways. It should also be relevant to *current* policy-making imperatives - ie driven by decision-making needs and therefore tailored to specific audiences.

Information should be simple, timely, policy-relevant and delivered in recognised ways.

6.2 INFORMATION PRODUCTS

6.2.1 Overview

This chapter examines the development of *information products*, rather than the systems necessary to deliver them. Figure 14 illustrates the life history of an information product in the form of an 'information pyramid'. Evidently, the transition from primary data to information product is one of integration, analysis and publishing.

As we saw in Chapter 2, different agencies take responsibility for different stages of this process. Primary data are collected and stored by custodian agencies who, upon request, send reported versions of their data to the agency or unit acting as the information system 'hub' (see Figure 3). The latter integrates the incoming reports, analyses them, and publishes the results in the form of information products. These are then delivered to selected audiences who are responsible for taking decisions on the issues

Figure 14: Information production pyramid (adapted from Hammond *et al* 1995)

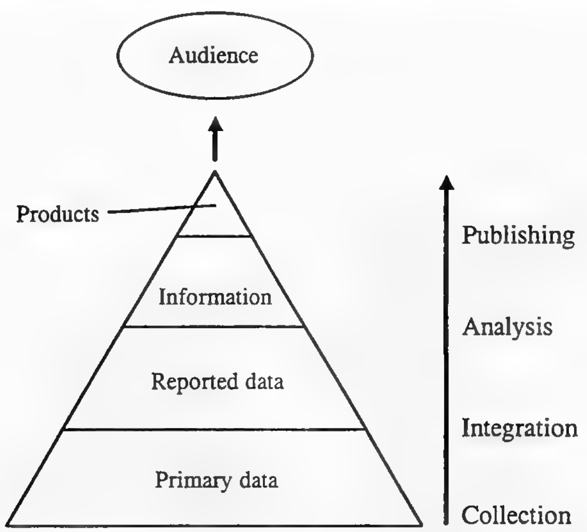
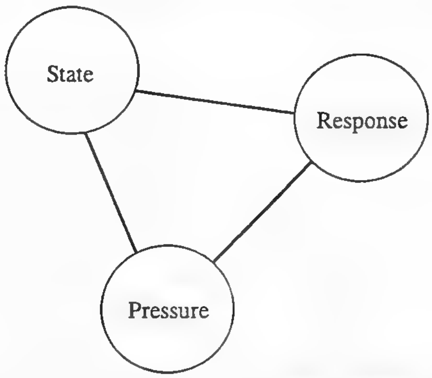


Figure 15: Elements of biodiversity information



concerned. Of course, custodian agencies are free to develop their own information products without reference to the hub. However, the latter is essential for creating collaborative products based on data from multiple agencies.

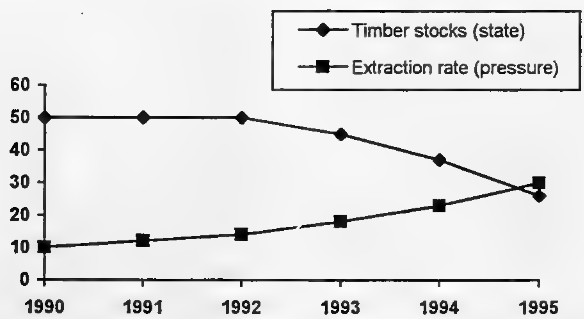
The key issues in information production are product content, data integration, data analysis and publishing.

6.2.2 Product Content

Having introduced a process for developing information products, what should the latter contain? The answer to this question naturally depends on which hazard, benefit or other environmental situation the product focuses on. Nevertheless, although products may differ significantly between countries, situations and groups, some generic suggestions for content can be made in the light of on-going research on indicators (for further information on the use of environmental indicators see Hammond *et al* 1995).

Biodiversity information can be divided into three essential elements: information on the *state* of the biological resource(s), information on the *pressure(s)* (both human-induced and natural) being applied to the resource(s), and information on the management *response(s)* being undertaken (see Figure 15). These tightly inter-linked

Figure 16: Example state and pressure indicators



elements correspond to 'what is happening?', 'why is it happening?', and 'what are we doing about it?' (Hammond *et al* 1995).

The usefulness of these elements to decision-making depends on local factors such as the prevailing legislation, the capacity of resource management agencies to act, and the complexity of the problems highlighted. However, it is generally accepted that information on pressure is the most policy-relevant form of information (A.Hammond, pers. comm.).

As an example, imagine an information product is developed to illustrate the diminishing stock of timber in a forest reserve. A graph showing the decline in timber volume (a 'state' indicator) year by year would certainly be useful; but a graph showing timber extraction rates (or another 'pressure' indicator) over the same period is more revealing, since this is more suggestive of a policy response (see Figure 16).

To help clarify what responses are necessary to combat environmental pressures, trends can be annotated with performance targets and thresholds, beyond which conditions are unsustainable or hazardous. Figure 17 illustrates the addition of a performance target for timber extraction rate, which if achieved, ensures the practice is sustainable. Figure 18 illustrates the use of a danger threshold, for example the density of an introduced species of river weed, beyond which access to fishing grounds is prevented.

Information products should embrace three elements of an environmental situation - state, pressure and response; in general, information on pressure is most critical for decision-making.

6.3 DATA INTEGRATION

6.3.1 Overview

To adequately inform on all elements of an environmental issue, data may need to be gathered from a wide range of sources. The successful integration of these data is a key factor in determining the success of the final product.

Data occur in a variety of *formats*, *media* and *types*, all of which introduce differences between datasets which potentially impede integration. So common is the integration problem that a whole industry has grown up to provide solutions in the form of data conversion and integration tools. However, tools alone cannot be relied upon to offer integration solutions. More importantly, data should be managed in ways which inherently

Figure 17: Example of a sustainability target

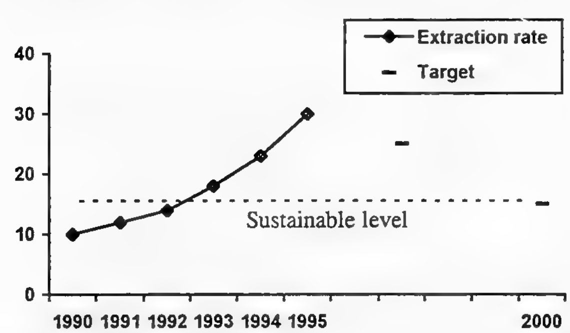
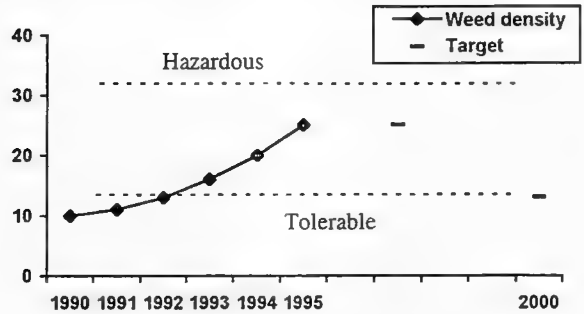


Figure 18: Example of a 'danger' threshold



facilitate integration (see Section 4.1).

Ideally, data are stored in such a way that integration is implicit, allowing product developers to concentrate their efforts on analysis and presentation, rather than data conversion and manipulation.

Rather than rely on the use of data conversion and integration tools, data should be managed in ways which inherently facilitate integration.

6.3.2 Data Format and Media

Integration problems caused by differing formats and media can normally be resolved via the use of standard computer equipment and data conversion software. For instance, if a dataset containing images of global forest cover is made available to national agencies, the provider will indicate the format and media of the dataset. Potential formats include Macintosh, DOS, Windows, UNIX (which signal different operating systems), plus details of any application software or hardware requirements; potential media include floppy disk, magnetic tape, CD-ROM, or even hard copy. If the national agency does not possess the necessary equipment there are two ways forward: the first is to acquire new equipment, which may or may not be cost-effective, depending on the value of the dataset; the second is to request an alternative format and/or media from the provider.

6.3.3 Data Type

Integrating different *types* of data is frequently more difficult than integrating data in different format and media. The reason is that fewer tools exist for direct conversion between data types; the technology which is available more likely provides linkages (bridges) between different data types. Some common types of biodiversity data are described below.

- Tabular Data

  Tabular data can be divided into numeric and categoric types:

  1. Numeric data are derived directly from many types of survey work, ranging from counts of species, to measurements of rainfall, tree growth or the length of a bird's primary feathers (which might be used in identification and taxonomic work). Numeric data can also be generated automatically from climatic recording machines, or derived from remotely-sensed images. Numeric data lends itself to computer-aided analysis, and the derivation of further datasets based on such analyses. For example, the absolute altitudinal range of a protected

area can be derived from subtracting the lower altitude from the upper. Such data are also extensively used in modelling. For example, information on the temperature, rainfall and altitude of a particular site (all numeric data) can be used to predict the Holdridge life zone within which it lies. It is possible to structure numeric data very strictly and exercise stringent validation procedures during data entry. Numeric data are easily stored as tables within database management systems, spreadsheet programs, statistics packages and so on. Key types of product which may be derived from numeric data include indices, tables, charts, graphs, and thematic maps.

2. Categoric data frequently occur in the environment. Examples include classified or coded non-numeric data, such as descriptions of soil type, land cover, forest type, life form, protected area designation, and so on. The data are usually structured through a thesaurus or data dictionary, and can be restricted to allowed values. Although statistical analysis may not be appropriate, categoric data are frequently used for database searches. For instance, if a life form category was given to every plant distribution record in a database, it would be simple to list all the 'tree' records, provided 'tree' was a life form category. Like numeric data, categoric data are also easily stored in tabular form since they are highly structured and contain a fixed set of values. However, the latter require careful definition, since changes in information needs can result in existing categories becoming obsolete (see Section 4.2). Typical products based on categoric data include tables, charts, and thematic maps.

- Textual Data

  Text is by far the most common type of biodiversity data. Examples include descriptions of protected areas, ecosystems, pressures and threats, or 'State of the Environment' reports, legislation, regulation, strategies and plans. By comparison with tabular data, text is much less structured, often subjective, poorly standardised, and difficult to analyse and maintain. For this reason, it is usually stored in word-processor format on computers, rather than a more formal data structure. Integration of textual and tabular data sources is a common problem. When combined with other forms of data (for instance tabular data presented in charts or maps), text is extremely valuable in setting context, presenting conclusions, and providing supporting information such as quality assessment or acknowledgements.

- Spatial Data

  Spatial data are playing an increasingly important role in biodiversity information products, since they effectively represent patterns and processes in the environment around us. Examples include point location records for species, species ranges, protected area boundaries, plus of course baseline geographic and biogeographic phenomena such as climate, topography, vegetation, administrative boundaries, land cover and land use. They may be maintained on paper maps, held in remotely-sensed digital format, or in computer-based geographic information systems. Like text, the integration of spatial data with other types is also a challenge. Typical products resulting from spatial data include all kinds of map, charts, graphs (eg of predicted habitat extent), and indices.

- Graphics, Video and Sound

  Photographs, diagrams, images and other visual materials are collectively referred to as graphics. They are a common form of report enhancement, often conveying ideas far more succinctly than with text. Hard copy graphics can be converted into computer graphics files by a procedure termed 'scanning', or recreated in the computer from scratch using graphics software. Moving images, such as video sequences or animations, and sound recordings, can also be incorporated into information products, particularly multimedia products (see 'Multimedia' in Section 6.5.3).

6.3.4 Examples

6.3.4.1 Text and Tabular Data

Although small amounts of text can be stored in database tables (in character and memo fields), it remains unformatted, without font changes, *italics* or **bold**. This is gradually being overcome as software manufacturers agree on connectivity protocols. For instance, the latest data management packages permit documents to be embedded as 'objects' in special fields. Less sophisticated solutions include the establishment of fixed links to external word processing packages via pointers stored in database fields, and the use of internal word processors within the database application.

One method of integrating text and tabular data is to import the required data directly into the text, avoiding the need to build any kind of formal bridge. This method is capable of achieving quick and pleasing results, since the word-processing application used to store the text may have sophisticated layout features. However, whilst the method is ideal for creating publications (see Section 6.5.3), it is not a surrogate for data management, since the imported data are removed from their parent database and cannot be kept up to date (the integration process must be repeated each time the parent database changes).

Documents can be embedded as 'objects' in modern data management packages. In the reverse direction, tabular data may be imported into documents to create publications.

6.3.4.2 Spatial and Tabular Data

Most biodiversity data relate in some way to specific geographic locations. For instance, protected areas have geographic boundaries, species have distributions, human and natural forces (eg rainfall) have geographic zones of influence. Data relating to geographic features may be stored in tabular form in database or spreadsheet applications. However, such applications normally have no facilities to respond to spatial enquiries, such as 'is this site within this region?', or 'how many hectares of this vegetation type occur at an altitude of less than 200 metres?'.

Spatial analyses of this kind are achieved through the use of geographic information systems (GIS) and desktop mapping packages. These maintain tabular data and spatial data components, the former being referred to as 'attributes' of geographic features. In order to link externally held tabular data into the map, a common field or identifier must exist in both table and geographic attribute file, enabling the tabular data to be associated with particular spatial features, such as sites.

As an example, suppose a series of numeric values are entered into a database table. The numbers refer to soil toxicity levels on successive months at a site in a highly mechanised agricultural area. A map of the study region exists in a desktop mapping program, and the task is to display the numeric data as a graph at the correct location on the map. This is a classic integration problem involving two types of data: tabular numeric and spatial.

A common field held in both database table and map might be the site code or geo-reference of the monitoring station where the toxicity was recorded. Having established that this field is common to both table and map, the mapping program can be requested to form the desired link allowing the toxicity values in the table to be fetched for purposes of display. For example, a particular colour or symbol could be applied to the site marker indicating its level of toxicity.

Integration of tabular and spatial data can be achieved by linking the two together via a common field such as a site code or geo-reference.

6.4 DATA ANALYSIS

6.4.1 Overview

The key to good data management is to manage data in such a way that varied analyses can be performed without the need for constant modification. Data analysis techniques empower the product designer to expose and summarise the key features of an environmental situation, in quantitative or qualitative ways. The end result is the production of one or more 'indicators', such as statistics, charts, or maps, which may be interpreted easily by decision-makers and lead to appropriate actions being taken.

6.4.2 Levels of Analysis

Elementary data analysis procedures, such as summation and averaging, are standard features of most data management software. Given suitably managed data, simple calculations can be performed such as the total number of wild crop varieties recorded in a particular valley, or the average abundance level of a species nation-wide. Results can often be summarised in the form of numbers, tables, graphs and charts.

However, in some situations it is necessary to apply complex, possibly spatial analyses to biodiversity data in order to obtain the desired indicator. Examples of situations demanding more complex analyses are:

- assessment of trends in space or time, for instance the depletion of resources in a buffer zone (time-series analysis)
- assessment of habitat suitability for different groups, eg endemic crop varieties (canonical analysis, pattern recognition)
- assessment of the degree to which protected areas adequately represent nationally available ecosystems, species, or genetic resources (clustering techniques, 'complementarity' metrics)
- classification of land use or vegetation types from remotely-sensed imagery (image processing, pattern recognition)
- environmental impact assessment.

There are two basic approaches to more complex data analysis: the use of packages (commercial or academic) containing pre-defined or customisable routines; and custom program design, in which the analysis routines are written from scratch.

Simple indicators may be developed using the elementary statistical facilities of popular data management packages; more complex analyses require specialist packaged software or custom programming techniques.

6.4.3 Packages

The first approach to complex data analysis involves the use of commercial (or academic) packages. These can be divided into the following groups:

- statistics packages
- GIS and desk-top mapping packages
- image analysis packages
- modelling packages
- expert systems and decision-support tools.

The use of packaged software can greatly reduce implementation time compared with custom programming and, more importantly, improve compatibility with other institutions for data exchange. Indeed, it is worth establishing an informal policy to limit the number of different packages in use by co-operating institutions to a short list, to take full advantage of the built-in compatibility and shared support which this will bring.

A disadvantage of packaged software is that the user is constrained to the functions included in the software, that is, it is usually not possible to add an additional function or operation at a later date. However, this need not be a problem if a good range of options are provided, and thus it is important to select the software carefully. Useful criteria to consider include:

- compatibility with existing configuration (hardware and operating system)
- compatibility with existing software
- richness of functions of the package
- peer expertise
- popularity in other institutions for similar tasks
- quality of technical support (eg documentation, locally available staff, telephone hotline, newsletter).

Information on some common packages is provided in UNEP/WCMC (1995).

Packaged analysis software can be acquired to develop complex indicators. If possible, the chosen package should be compatible with existing hardware and software.

6.4.4 Custom Program Design

Data analysts who possess a good knowledge of statistical theory and programming concepts can write 'custom' analysis routines using a computer programming language. Options include the macro language of the DBMS or spreadsheet package managing the data, or a high-level language such as BASIC, FORTRAN, C, or PASCAL. In some cases the task may be simplified by drawing on third party 'libraries' of commonly used statistical routines or, alternatively, implementing published program listings or 'numeric recipes' directly.

An example of this approach might be the calculation of an economic value index for a series of managed areas, which might involve totalling economically valuable species in each area, weighting each species according to its particular economic value. An analysis of this kind would require, perhaps, a 20-30 line program, provided access to all the necessary data was straightforward.

Very complex analyses can be undertaken using custom programming techniques; indeed much academic research is conducted in this way. Examples include modelling the impact of climate change on natural and managed ecosystems, modelling the effect of forest management practices on tree regeneration, assessing the effects of population growth on ecosystem health, and calculating the wilderness value of different landscapes.

Occasionally, very good routines are released into the public domain by academics, government researchers, conservationists, and others directly involved in biodiversity research. However, whilst it is preferable to make use of existing software where possible, it should be recognised that such programs will nearly always require modification to suit local or national analysis needs. Useful factors to bear in mind before employing such programs are:

- suitability to local conditions
- scientific peer acceptance
- compatibility with existing applications
- quality of technical support.

Many of the areas requiring custom programming work are at the cutting edge of current knowledge, and are thus evolving rapidly. For this reason most available biodiversity-related models apply only to specific local situations and may only be valid elsewhere under restricted circumstances. More generic models have been developed in more mature disciplines such as agriculture, forestry, and meteorology. Programs modelling national biodiversity sustainability are at a very early stage of research.

Information on some existing programs and modelling software is given in UNEP/WCMC (1995).

Sophisticated indicators can be developed using custom programming techniques. However, when applying programs developed by others, attention should be paid to local suitability and peer acceptance.

6.5 PUBLISHING INFORMATION

6.5.1 Overview

The final stage in the development of an information product is packaging, communication and marketing (awareness-raising) - activities which are collectively known as publishing. Without attention to this crucial, but often neglected stage, the impact of the product will be lower. For instance it is unlikely that commercial publishers would neglect to market their books or journals; they know that unless a product is attractive, easily available, and clearly promotes its content, its audience will be small. After all, there is an abundance (some say too many) of publicly available information sources: why should they choose yours?

Without sufficient attention to information packaging, communication and marketing, the impact of an information product will be lower.

6.5.2 Traditional Publications

Information products are traditionally prepared as reports, papers, pamphlets, brochures and other forms of publication, including video, audio, slides and posters for large audiences. Some distinguishing characteristics of good publications are described below:

- Structure

  Logical flow to the information, with a well defined beginning, middle and end; often a brief summary at the very start (for example an 'executive summary' to a report); detail consigned to annexes, less visible areas, or completely left out; efficient 'navigation' aids for large publications, such as table of contents, list of figures, page numbering, references, and index.

- Layout

  Not overcrowded - just clear, simple information delivery; plenty of space and features surrounding the body of the message; attractively designed pages with clear route through the information; judicious use of shading, colour, and fonts; diagrams, tables, maps, charts, graphs, photographs, images, and other 'features', to enhance key messages and break up text; boxes containing summaries or supplementary text; examples and case studies to reinforce key points.

- Access

  Efficient mechanism for publication delivery, free from burdensome procedures.

- Cost

  Available at a cost which is affordable by the target audience and sustainable indefinitely by the providing organisation, in terms of time, money and administrative overheads.

- Quality

  Uses the best available scientific knowledge; intermediate publications, data sources, and intellectual property used are fully disclosed; clear copyright, ownership, and follow-up details.

Good publications meet certain structural, layout, access, cost and quality standards.

6.5.3 Electronic Publications

In much the same way that text, images, and tabular data can be arranged together in a report or book, computer software exists to integrate these data, plus other types of data such as sound and video. In general, the same characteristics which enhance a traditional publication apply equally to electronic publications.

The production of electronic publications should not be confused with electronic data management. Publications consist of selected pieces of information originating, wherever possible, from well managed data sources. Once incorporated in a publication, the information becomes out of date as soon as the parent data changes. Thus publications convey information based on a 'snapshot' of the latest data, but are not a surrogate for on-going data management.

Various kinds of electronic publication are described below. Each requires an increasing degree of investment in computer hardware and software to fabricate.

- Document Viewers

  The simplest form of publication is the word-processing document, in which text usually dominates, but images and tabular data may be added. Once prepared, word-processing documents can be disseminated by various means, including hard copy, floppy disk, and on-line transfer. However, the recipient of the document must have a copy of the word-processing package to view it, or at minimum have a similar package which can import and display the document.

  Recently, some manufacturers have developed free software to enable users to view (but not edit) word-processing documents without the need for an existing package. An example is Microsoft, who distribute a program called 'WordView' to enable users to view Microsoft Word documents. This program can be distributed freely with Microsoft Word files to enable users to view their contents - a simple and straightforward means of preparing an electronic publication.

- Slide Presentations

  Purpose-built computer presentation software (eg Microsoft PowerPoint) enables the developer to construct a series of screen-sized 'slides' which, when displayed one after another, produce a professional looking presentation. Most presentation software allows the presenter to run through the slide show under manual control (eg by clicking the mouse), or under automatic control, in which case the computer changes slide after a predefined interval. When presenting to large audiences, the computer can be connected to a device known as a projector panel which projects the computer screen on to a wall (this technique is useful for many types of computer presentation).

  The slides themselves can be furnished with all kinds of text and graphics, and links can be made within the slides to additional computer demonstrations such as video sequences or sound recordings.

- Document Management Software

  Document management software are available which offer additional features to most word-processing packages. These features include:

- ✓ Hypertext - the ability to embed 'hyperlinks' in the document which allow the reader to 'navigate' to and from different sections of the document, or indeed to other documents, by clicking a computer mouse (eg Folio Views, Adobe Acrobat, I-View, Netscape)
- ✓ Fully indexed search - which permits rapid and sophisticated text searches to be conducted over very large documents (eg Folio Views, Adobe Acrobat)
- ✓ Cross-platform portability - the ability to share or exchange documents in a single format which is readable by almost any computer, independent of its specification (eg Adobe Acrobat).

All these features are useful in certain situations. For example, WCMC collaborated with the Indira Gandhi Conservation Monitoring Centre (IGCMC) in India, to develop a hypertext 'Biodiversity Profile of India'. This product contained a mixture of text and graphics, including a series of maps of forest cover, endemic bird areas and protected areas. So as not to disrupt the flow of the text, the maps were made accessible to the reader via hyperlinks. The product was developed using an off-line hypertext development tool based around the HTML standard (see 'On-Line Publishing' below), allowing it to be released over the Internet without further modification.

Sophisticated document management software, such as Folio Views and Adobe Acrobat, offer complete systems for electronic document delivery. They are most appropriate when sophisticated hypertext or text-searching facilities are required. For example, the *Electronic Resource Inventory* (UNEP/WCMC 1995) was created using Folio Views.

- Multimedia

Multimedia products offer seamless integration of text, graphics, sound, video and animation. By clicking the mouse on relevant text or graphics, the user can select different screens of information or multimedia displays. True multimedia is constrained only by the imagination of the developer, encouraging the development of novel and exciting ways of revealing information to users.

Although absorbing to use, multimedia products are beyond the realm of most developers, since typical products cost huge sums of money to make. The following reasons account for this:

- ✓ highly qualified artists and developers may be required
- ✓ the product may take months or even years to complete
- ✓ much research may be needed to obtain video footage, sound recordings, or other multimedia features
- ✓ costs may be incurred in licensing other intellectual property.

- On-Line Publishing

Hypertext systems have recently become popularised by the emergence of the World Wide Web (WWW) as the premier tool for viewing information on the Internet - the publicly accessible global communications network. Access to the WWW ('Web' for short) is achieved using hypertext software such as Netscape and Mosaic, which may be downloaded from their suppliers. Such items of software, which are frequently referred to as 'Web browsers', can be used without any prior knowledge of computers.

To publish information on the Internet, or via any 'on-line' communications service, information must be permanently (or at least near-permanently) accessible by users. This requires dedicated computer equipment and a dedicated communications path. Once this investment has been made, the process of releasing documents over the Web is relatively straightforward. First a 'home page', or starting point for on-line users is created. Then hyperlinks to other documents and data sources are added. All pages of information are formatted according to the HTML (HyperText Markup Language) standard, which is easy to learn yet powerful enough to create imaginative and highly structured pages.

An impressive 'home page' is offered by the Australian Environmental Resources Information Network (ERIN) at http://www.erin.gov.au. From this page, a wide variety of environmental information may be accessed by decision-makers and the general public alike. A feature of the ERIN presentation is the ability to perform custom database searches and create custom maps, all via the same intuitive interface. This is achieved by linking the presentation with expertly managed tabular and spatial datasets via a standard information exchange protocol such as SQL.

An example of this process is a hypertext page describing a particular bird species, say the African Shikra. On the page is a form inviting the browser to select a particular geographic region in Africa. Upon request, a query is sent out to a database of distribution records, and any which occur within the desired region are returned as a new hypertext page.

A range of electronic publication techniques are available, ranging from document viewers to multimedia development systems and on-line publishing. The decision which to chose depends on the ability to invest in the necessary computer hardware, software, communications and human resources.

7

Supporting Materials

7.1 REFERENCES

Archer, H. and Croswell, P.L. 1989. Public access to geographic information systems: an emerging legal issue. *Photogrammetric Engineering and Remote Sensing*. **55**:1575-1581.

Aronoff, S. 1989. *Geographic Information Systems - A Management Perspective*. WDL Publications, Ottawa, Canada.

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

Ayers, L.F., Kottman C.A. 1994. A call for GIS certification. *GIS World* **7** (12):48-52.

British Standards Institute 1994. *Draft BS ISO 14001: Environmental Management Systems - Specification with guidance for use*. Technical Committee ESS/1 - Environmental Management Systems, British Standards Institute, London, UK.

Burnett, J., Copp, C. and Harding, P. 1995. *Biological Recording in the United Kingdom: Present practice and future development*. Summary report. Department of the Environment, London, UK.

Busby, J.R. and Walton, D.W. 1994. *A National Biological Survey for the United States? Comparable Australian Activities at the National Level*. In: Biodiversity - Broadening the Debate. Longmore, R. (Editor). Australian Nature Conservation Agency, Canberra.

Burley, C. 1994. *CIESIN Metadata Entry Form Instructions*. CIESIN.

Chen, P.P. 1976. *The entity-relationship model - toward a unified view of data. ACM Trans. on Database Systems* **1**(1):9-36.

Clark, G.L. 1981. Law, the state and the spatial integration of the United States. *Environment and Planning*. **A13**(10):1197-1232.

Codd, E.F. 1970. A relational model of data for large shared data banks. *Comm. ACM*. 13(6):377-387.

Codd, E.F. 1979. Extending the database relational model to capture more meaning. *ACM Trans. Database Systems* **4**(4):397-434.

Connell, J.L., Shatner L.B. 1989. *Structured Rapid Prototyping: An Evolutionary Approach to Software Development*. Prentice-Hall, Englewood Cliffs, NJ.

Cooley, G.P., et al. 1992. *Collections and Research Information System Master Plan*. MTR_92W00083 Mitre, McLean, Virginia, USA.

Cooley G.P., Harrington, M.B., and Lawrence, L.M. 1993. *Analysis and Recommendations for Scientific Computing and Collections Information Management of Free-Standing Museums of Natural History and Botanical Gardens*. MTR-93W0000109V1 Vol.1. Mitre, McLean, Virginia, USA.

Crain, I.K. 1992. *User requirements for the Harmonization of Environmental Measurement Information System (HEMIS)*. UNEP-HEM Munich, Germany. 86pp.

Crain, I.K. (Editor) 1994. *An Introduction to HEM and the HEMDisk*. UNEP-HEM Munich, Germany.

Cutts, G. 1991. *Structured Systems Analysis and Design Methodology*. 2nd edition. Blackwell Scientific, Oxford, UK. 481pp.

Date, C.J. 1983. *An Introduction to Database Systems*. Vol. II. Addison-Wesley, Reading, Mass. 383pp.

Date, C.J. 1990. *An Introduction to Database Systems*. 5th edition., Vol. I. Addison-Wesley, Reading, Mass.

DeMarco, T. 1979. *Structured Analysis and System Specification*. Prentice-Hall, Englewood Cliffs, NJ. 352pp.

Edwards, P. 1983. *Systems Analysis and Design*. Mitchell McGraw Hill.

Epstein, E.F. 1990. Access to Information: Legal Issues, *Proceedings of the XIX Congress of the International Federation of Surveyors(FIG)*. Vol 3. 92-99pp.

Fidel, R. 1987. *Database Design for Information Retrieval*. John Wiley, New York. 232pp.

Fitzgerald, G., Stokes, N., and Wood, J.R.G. 1985. *Feature analysis of contemporary information system methodologies*. Computer Journal **28**(3):223-230.

Flaaten, P. 1989. *Foundations of Business Systems*. Dryden Press.

Gane, C. 1990. *Computer Aided Software Engineering: The Methodologies, the Products and the Future*. Prentice-Hall, Englewood Cliffs, NJ. 220pp.

Gause, D.C and Weinberg, G.M. 1989. *Exploring Requirements: Quality Before Design*. Dorset House Publishing Company.

Hammond, A., et al. 1995. *Environmental Indicators: A Systematic Approach to Measuring and Reporting on Environmental Policy Performance in the Context of Sustainable Development*. World Resources Institute, Washington.

Howe, D.R. 1983. *Data Analysis for Data base Design*. Arnold, London. 307pp.

Jordan, E., and Machesky, J. 1990. *Systems Development*. PWS-Kent.

Kroenke, D.M. 1992. *Database Processing*. Macmillan.

Maddison, R.N., et al. 1983. *Information System Methodologies*. Wiley Heyden, Chichester, UK. 128pp.

Malamud, C. 1989. *INGRES: Tools for Building an Information Architecture*. Van Nostrand Reinhold.

Margules, C.R. and Redhead, T.D. 1995. *Guidelines for using the BioRap Methodology and Tools*. CSIRO, Australia, 1995.

McLean, I. 1989. *Democracy and New Technology*. Polity Press, Cambridge. 204pp.

MUIENR/WCMC 1995. *Biodiversity Data Bank: Software Overview*. Makerere University Institute of Environment and Natural Resources, Kampala, Uganda, 1995.

Obermeyer, N.J., and Pinto, J.K. 1994. *Managing Geographic Information Systems*. The Guilford Press. 226pp.

Olle, T.W., Sol, H.G., and Verrijn Stuart A.A. (Editors). 1982. *Information Systems Design Methodologies - A Comparative Review*. North Holland, Amsterdam.

Olle, T.W., Sol, H.G., and Tully, C.J. (Editors). 1983. *Information Systems Design Methodologies - A Feature Analysis*. North Holland, Amsterdam.

Onsrud, H.J. 1989. Legal and Liability Issues in Publicly Accessible Land Information Systems, *Proc. GIS/LIS*, Vol.1. 295-300pp.

Oxborrow, E. 1989. *Databases and Database Systems*. Chartwell-Bratt, Bromley, UK. 254pp.

Pinborg, U. 1992. *Catalogue of Data Sources (CDS) for the Environment: Analysis and Suggestions for a Meta-data System and Service*. European Environment Agency.

Powers, M.J. and Cheney, P.H. 1990. *Structured Systems Development*. Boyd and Fraser Publishing.

Rhoads, A.F. 1990. *A modern, computer-accessed Flora of Pennsylvania: a tool for resource managers*. In: [Ecosystem Management. New York State Museum, New York].

Richardson, B.J. 1994. The industrialisation of scientific information. In: Forey, P.L., Humphries, C.J., and Vane-Wright, R.I. (Editors) Systematics and Conservation Evaluation. *Systematics Association Special Volume* **50**:123-31. Clarendon Press, Oxford, UK.

Robinson, H. 1981. *Database Analysis and Design*. Chartwell-Bratt, England. 375pp.

Rock-Evans, R. 1981. *Data Analysis*. IPC Business press.

Smith, J.M., and Smith, D.C.P. 1977. Database Abstractions: Aggregation and Generalization. *ACM Trans. Database Systems* **2**(2):105-133.

Stein, B.A. 1994. *Strengthening National Capacities for Biodiversity Information Management*. The Nature Conservancy, USA.

Townsend, J.T. 1992. *Introduction to Databases*. Que.

Ullman, J.D. 1982. *Principles of Database Systems*. 2nd edition. Computer Science Press, Rockville, Maryland. 484pp.

INFOTERRA 1990. *Thesaurus of Environmental Terms*. 3rd edition. UNEP, Nairobi.

UNEP 1992. *The Grid Meta-Database (MDb) Entity-Attribute Definitions*. UNEP-GRID Geneva.

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

UNEP/WCMC 1995. *Electronic Resource Inventory: A searchable resource for biodiversity data management*. WCMC, Cambridge, UK. [Windows 3.1 or higher].

Verheijen, G.M.A., and Bekkum, J.van 1982. NIAM: An Information Analysis Method, In: Olle, T.W., et al (Editors), *Information Systems Design Methodologies - A Comparative Review*. North-Holland, Amsterdam.

Vinden R.J. 1982. *Data Dictionaries for Database Administrators*. TAB Books, Blue Ridge Summit, PA. 176pp.

Van Dijkhuizen, H. 1994. *World Bird Database: User Requirement Specification and System Design Specification*. BirdLife International, Cambridge, UK.

Walter, K.S. 1989. Designing a computer-software application to meet the plant-record needs of the Arnold Arboretum. *Arnoldia* **49**(1):42-53.

WCMC 1993. *Availability of Biodiversity Information in East Africa*. WCMC, Cambridge, UK.

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

WDC 1991. *Directory Interchange Format Manual (version 4.0)*. World Data Center, NASA, USA.

Wiggins, L.L. and French, S.P. 1992. *Geographic Information Systems: Assessing Your Needs and Choosing a System*. Planning Advisory Service Report. American Planning Association, Chicago.

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

WRI/UNDP 1990. *World Resources 1990-92*. Oxford University Press, New York. pp13-19.

WRI/UNEP/IUCN 1995. *National Biodiversity Planning - Guidelines Based on Early Experiences Around the World*. World Resources Institute, Washington.

WWF-India 1994. *Indira Gandhi Conservation Monitoring Centre: A Profile*. WWF-India, New Delhi.

Yourdon, E. 1975. *Techniques of Program Structure and Design*. Prentice-Hall, Englewood Cliffs, NJ.

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

Yourdon, E. Undated. *The CASE Report*. Nastec Corporation, Southfield, MI, USA.

7.2 GLOSSARY

7.2.1 Biodiversity Terms

Accession. A sample of a crop variety collected at a specific location and time; may be of any size.

Alien species. A species occurring in an area outside of its historically known natural range as a result of intentional or accidental dispersal by human activities (also known as an exotic or introduced species).

Artificial insemination. A breeding technique, commonly used in domestic animals, in which semen is introduced into the female reproductive tract by artificial means.

Assemblage. See 'Community.'

Biochemical analysis. The analysis of proteins or DNA using various techniques, including electrophoretic testing and restriction fragment length polymorphism analysis. These techniques are useful methods for assessing plant diversity and have also been used to identify many strains of micro-organisms.

Biodiversity. See 'Biological diversity'.

Biogeography. A branch of geography that deals with the geographical distribution of animals and plants.

Biological diversity. Means the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part;. this includes diversity within species, between species and of ecosystems.

Biological Oxygen Demand (BOD). The amount of dissolved oxygen consumed by micro-organisms as they decompose organic material in polluted water. Measurement of the rate of oxygen take-up is used as a standard test to detect the polluting capacity of effluent; the greater the BOD value (g) (and hence the greater the presence of oxygen - consuming micro-organisms) the greater the volume of pollutant present.

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

Biologically unique species. A species that is the only representative of an entire genus or family.

Biome. A major portion of the living environment of a particular region (such as a fir forest or grassland), characterised by its distinctive vegetation and maintained by local climatic conditions.

Bioregion (bioregional planning). A territory defined by a combination of biological, social, and geographic criteria, rather than geopolitical considerations; generally, a system of related, interconnected ecosystems.

Biosphere reserve. Established under UNESCO's Man in the Biosphere (MAB) Program, biosphere reserves are a series of protected areas intended to demonstrate the relationship between conservation and development.

Biota. The living organisms of a region.

Biotechnology. Techniques that use living organisms or substances from organisms to make or modify a product. The most recent advances in biotechnology involve the use of recombinant DNA techniques and other sophisticated tools to harness and manipulate genetic materials.

Biotic. Pertaining to any aspect of life, especially to characteristics of entire populations or ecosystems.

Breed. A group of animals or plants related by descent from common ancestors and visibly similar in most characteristics. Taxonomically, a species can have numerous breeds.

Breeding line. Genetic lines of particular significance to plant or animal breeders that provide the basis for modern varieties.

Buffer zone. The region near the border of a protected area; a transition zone between areas managed for different objectives.

Captive breeding. The propagation or preservation of animals outside their natural habitat, involving control by humans of the animals chosen to constitute a population and of mating choices within that population.

Carrying capacity. The maximum number of people, or individuals of a particular species, that a given part of the environment can maintain indefinitely.

Chromatography. A chemical analysis technique whereby an extract of compounds is separated by allowing it to migrate over or through an adsorbent (such as clay or paper) so that the compounds are distinguished as separate layers.

Climax community. The end of a sequence of successions; a community that has reached stability under a particular set of environmental conditions.

Clonal propagation. The multiplication of an organism by asexual means such that all progeny are genetically identical. In plants, it is achieved through use of cuttings or in vitro culture. For animals, embryo splitting is a method of clonal propagation.

Co-management. The sharing of authority, responsibility, and benefits between government and local communities in the management of natural resources.

Common property resource management. The management of a specific resource (such as a forest or pasture) by a well-defined group of resource users with the authority to regulate its use by members and outsiders.

Community. A group of ecologically related populations of various species of organisms occurring in a particular place and time.

Comparative advantage. Relative superiority with which a region or state may produce a good or service.

Complementarity. The concept of achieving conservation efficiently by ensuring that a set of areas is assembled with due regard to the additional species that each brings into the network. This is the basis of a critical faunas analysis.

Conservation. The management of human interactions with genes, species, and ecosystems so as to provide the maximum benefit to the present generation while maintaining their potential to meet the needs and aspirations of future generations; encompasses elements of saving, studying, and using biodiversity.

Country of origin of genetic resources. Means the country which possesses those genetic resources in in-situ conditions.

Country providing genetic resources. Means the country supplying genetic resources collected from in-situ sources, including populations of both wild and domesticated species, or taken from ex-situ sources, which may or may not have originated in that country.

Critical faunas analysis. Is a methodology to identify the minimum set of areas which would contain at least one viable population of every species in a given animal or plant group.

Critical habitat. A technical classification of areas in the United States that refers to habitats essential for the conservation of endangered or threatened species. The term may be used to designate portions of habitat areas, the entire area, or even areas outside the current range of the species.

Cryogenic storage. The preservation of seeds, semen, embryos, or micro-organisms at very low temperatures, below -130°C . At these temperatures, water is absent, molecular kinetic energy is low, diffusion is virtually nil, and storage potential is expected to be extremely long.

Cryopreservation. See 'Cryogenic storage'.

Cultivar. A cultivated variety (genetic strain) of a domesticated crop plant (derived from 'cultivated variety').

Cultural diversity. Variety or multiformity of human social structures, belief systems, and strategies for adapting to situations in different parts of the world.

Cutting. Plant piece (stem, leaf, or root) removed from a parent plant that is capable of developing into a new plant.

Cycad. Any of an order of gymnosperms of the family cycadaceae. Cycads are tropical plants that resemble palms but reproduce by means of spermatozoids.

DNA. Deoxyribonucleic acid. The nucleic acid in chromosomes that codes for genetic information.

Domesticated or cultivated species. Means species in which the evolutionary process has been influenced by humans to meet their needs.

Domestication. The adaptation of an animal or plant to life in intimate association with and to the advantage of man.

Ecology. A branch of science concerned with the interrelationship of organisms and their environment.

Ecosystem. A dynamic complex of plant, animal, fungal, and micro-organism communities and their associated non-living environment interacting as an ecological unit.

Ecosystem diversity. The variety of ecosystems that occurs within a larger landscape, ranging from biome (the largest ecological unit) to micro-habitat.

Ecotourism. Travel undertaken to witness sites or regions of unique natural or ecological quality, or the provision of services to facilitate such travel.

Electrophoresis. Application of an electric field to a mixture of charged particles in a solution for the purpose of separating (eg mixture of proteins) as they migrate through a porous supporting medium of filter paper, cellulose acetate, or gel.

Embryo transfer. An animal breeding technique in which viable and healthy embryos are artificially transferred to recipient animals for normal gestation and delivery.

Endangered species. A technical definition used for classification in the United States referring to a species that is in danger of extinction throughout all or a significant portion of its range. The International Union for the Conservation of Nature and Natural Resources (IUCN) definition, used outside the United States, defines species as endangered if the factors causing their vulnerability or decline continue to operate.

Endemic. Restricted to a specified region or locality.

Endemic Bird Area (EBA). Is a term used by BirdLife International to describe areas with two or more restricted-range bird species entirely confined to them.

Endemism. The occurrence of a species in a particular locality or region.

Environmental Impact Assessment (EIA). A method of analysis which attempts to predict the repercussions of a proposed developments (usually industrial) upon the social and physical environment of the surrounding area.

Equilibrium theory. A theory of island biogeography maintaining that greater numbers of species are found on larger islands because the populations on smaller islands are more vulnerable to extinction. This theory can also be applied to terrestrial analogues such as forest patches in agricultural or suburban areas or nature reserves where it has become known as 'insular ecology.'

Exotic species. An organism that exists in the free state in an area but is not native to that area. Also refers to animals from outside the country in which they are held in captive or free-ranging populations.

Ex-situ. Pertaining to study or maintenance of an organism or groups of organisms away from the place where they naturally occur. Commonly associated with collections of plants and animals in storage facilities, botanic gardens or zoos

Ex-situ conservation. The conservation of components of biological diversity outside their natural habitats.

Extant. Species are those whose members are living at the present time.

Extinct. As defined by the IUCN, extinct taxa are species or other taxa that are no longer known to exist in the wild after repeated search of their type of locality and other locations where they were known or likely to have occurred.

Extinction. Disappearance of a taxonomic group of organisms from existence in all regions.

Fauna. Organisms of the animal kingdom.

Feral. A domesticated species that has adapted to existence in the wild state but remains distinct from other wild species. Examples are the wild horses and burros of the West and the wild goats and pigs of Hawaii.

Flora. Organisms of the plant kingdom

Forest Resource Accounting (FRA). A methodology for forest management based on the use of information for improved conservation and sustainable utilisation.

Gamete. The sperm or unfertilised egg of animals that transmit the parental genetic information to offspring. In plants, functionally equivalent structures are found in pollen and ovules.

Gene. A chemical unit of hereditary information that can be passed from one generation to another.

Gene bank. A facility established for the ex situ conservation of individuals (seeds), tissues, or reproductive cells of plants or animals.

General Circulation Model (GCM). Global-scale computer model that simulates physical and chemical processes in the atmosphere, both at the present time and in the future under conditions of elevated concentrations of radiatively active gases (enhanced greenhouse effect). In some instances integrated with comparable processes occurring at the surface and within oceans and at the land surface.

Genetic diversity. The variety of genes within a particular species, variety, or breed.

Genetic drift. A cumulative process involving the chance loss of some genes and the disproportion ate replication of others over successive generations in a small population, so that the frequencies of genes in the population is altered. The process can lead to a population that differs genetically and in appearance from the original population.

Genetic material. Means any material of plant, animal, microbial or other origin containing functional units of heredity.

Gene-pool. The collection of genes in an interbreeding population.

Genetic resources. Means genetic material of actual or potential value.

Genotype. The genetic constitution of an organism as distinguished from its physical appearance.

Genus. A category of biological classification ranking between the family and the species, comprising structurally or phylogenetically related species or an isolated species exhibiting unusual differentiation.

Germplasm. The genetic material, especially its specific molecular and chemical constitution, that compromises the inherited qualities of an organism.

Grassroots (organisations or movements). People or society at a local level, rather than at the centre of major political activity.

Grow-out (growing-out). The process of growing a plant for the purpose of producing fresh viable seed to evaluate its varietal characteristics.

Habitat. Is the environment in which an animal or plant lives, generally defined in terms of vegetation and physical features.

Hotspot. Is an area on earth with an unusual concentration of species, many of which are often endemic to the area.

Hybrid. An offspring of a cross between two genetically unlike individuals.

Hybridisation. Crossing of individuals from genetically different strains, populations, or species.

Important Bird Area (IBA). Sites of importance to birds, identified by BirdLife International and Wetlands International. The sites are identified for four groups of birds: regularly occurring migratory species which concentrate at and are dependent on particular sites either when breeding, or migration, or during the winter; globally threatened species (ie species at risk of total extinction); species and sub-species threatened throughout all or parts of their range but not globally; species that have relatively small total world ranges with important populations in specific areas.

In-situ. Maintenance or study of organisms within an organism's native environment.

In-situ conservation. The conservation of biodiversity within the evolutionary dynamic ecosystems of the original habitat or natural environment.

Inbreeding. Mating of close relatives resulting in increased genetic uniformity in the offspring.

Indicator species. A species whose status provides information on the overall condition of the ecosystem and of other species in that ecosystem.

Indigenous peoples. People whose ancestors once inhabited a place or country, and continue to live in conformity with their own social, economic, and cultural customs and traditions (also: 'native peoples' or 'tribal peoples')

Intellectual Property Rights (IPR). Rights intended to protect knowledge from being exploited without consent.

Inter-species. Between different species.

Intrinsic value. The value of creatures and plants independent of human recognition and estimation of their worth.

Introduced species. See 'Alien species'.

Inventory. On-site collection of data on natural resources and their properties.

In vitro. (Literally 'in glass'). The growing of cells, tissues, or organs in plastic vessels under sterile conditions on an artificially prepared medium.

Island biogeography. The study of the relationship between island area and species number. This idea has also been applied to isolated areas of habitat in continental areas which are effectively islands for many species. The extent to which habitat fragmentation may lead to extinction of species can be predicted from the relationship between number of species and island area.

Isoenzyme (Isozyne). The protein product of an individual gene and one of a group of such products with differing chemical structures but similar enzymatic function.

Keystone species. A species whose loss from an ecosystem would cause a greater than average change in other species populations or ecosystem processes.

Landrace. Primitive or antique variety usually associated with traditional agriculture. Often highly adapted to local conditions.

Land Mapping Unit (LMU). The smallest are of land that can be delineated on a map of a particular scale. Used in land evaluation as the basis of spatial variation.

Land Quality (LQ). A complex attribute of land, which acts in a manner distinct from the actions of other land qualities in its influence on the suitability of land for a specified kind of use.

Land Use Requirements (LUR). The requirements are related to growth and yield of crops and trees, animal husbandry, land management and conservation. The expression of the conditions for successful implementation are described for each LUT, eg growth requirements of certain tree species.

Land Utilisation Type (LUT). Described in terms of necessary inputs and expected results, based on a number of key attributes obtained from land use data; produce, capital input, labour input, farm size, land tenure, technical know-how, level of mechanism etc. LUTs relate to the physical social and economic conditions of the area and according to the development of objectives; description of the key attributes, reflecting biological, socio-economic and technical aspects of the production environment and which are relevant to the productive capacity of a LMU.

Living collections. A management system involving the use of off-site methods such as zoological parks, botanic gardens, arboretums, and captive breeding programs to protect and maintain biological diversity in plants, animals, and micro-organisms.

Marine Protected Area (MPA). An area of sea (or coast) especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means.

Megadiversity countries. Are the small number of countries, located largely in the tropics, which account for a high percentage of the world's biodiversity by virtue of containing very large numbers of species.

Micro-organisms. In practice, a diverse classification of all those organisms not classed as plants or animals, usually minute microscopic or submicroscopic and found in nearly all environments. Examples are bacteria, cyanobacteria (blue-green algae), mycoplasma, protozoa, fungi (including yeasts), and viruses.

Minimum Viable Population (MVP). The smallest isolated population having a good chance of surviving for a given number of years despite the foreseeable effects of demographic, environmental, and genetic events and natural catastrophes.

Minor breed. A livestock breed not generally found in commercial production.

Modelling. The use of mathematical and computer based simulations as a planning technique.

Morphology. A branch of biology that deals with form and structure of organisms.

Multiple use. An on-site management strategy that encourages an optimum mix of several uses on a parcel of land or water or by creating a mosaic of land or water parcels, each with a designated use within a larger geographic area.

Mycorrhizal fungi. A fungus living in a mutualistic association with plants and facilitating nutrient and water uptake.

National income accounts. System of record by which the vigour of a nation's economy is measured, (results are often listed as Gross National Product, or Gross Domestic Product).

Native. Indigenous to a particular locality or region.

Nitrogen fixation. A process whereby nitrogen fixing bacteria living in mutualistic associations with plants convert atmospheric nitrogen to nitrogen compounds that plants can utilise directly.

Non-Governmental Organisation (NGO). A non-profit group or association organised outside of governmental structures to realise particular objectives (such as environmental protection) or serve particular constituencies (such as indigenous peoples). NGO activities range from research, information distribution, training, local organisation, and community service to legal advocacy, lobbying for legislative change, and civil disobedience. NGOs range in size from small groups within a particular community to huge membership groups with a national or international scope.

Off-site. Propagation and preservation of plant, animal, and micro-organism species outside their natural habitat.

On-site. Preservation of species in their natural environment.

Open-pollinated. Plants that are pollinated by physical or biological agents (eg wind, insects) and without human intervention or control

Orthodox seeds. Seeds that are able to withstand the reductions in moisture and temperature necessary for long-term storage and remain viable.

Parataxonomists. Field-trained biodiversity collection and inventory specialists recruited from local areas.

Participatory Rural Appraisal (PRA). Also known as Rapid Rural Appraisal, PRA is a relatively new and different approach for conducting action-oriented research in developing countries. PRAs are used to help involve villagers and local officials leaders in all stages of development work, from the identification of needs and decision making to the assessment of completed projects. The term can be used to describe any new methodology which makes use of a multidisciplinary team.

Patent. A government grant of temporary monopoly rights on innovative processes or products.

Pathogen. A disease-causing micro-organism, bacterium or virus.

Phenotype. The observable appearance of an organism, as determined by environmental and genetic influences (in contrast to genotype).

Phytochemical. Chemicals found naturally in plants.

Phylogenetic. Pertaining to the evolutionary history of a particular group of organisms.

Phylum. In taxonomy, a high-level category just beneath the kingdom and above the class; a group of related, similar classes.

Population. A group of individuals with common ancestry that are much more likely to breed with one another than with individuals from another such group.

Population and Habitat Viability Assessment (PHVA). The theoretical modelling of minimum areas, habitat types and population sizes, to sustain any one or more species. Population size will be determined by the carrying capacity of the habitat.

Population Viability Analysis (PVA). The theoretical determination of the minimum viable (in terms of genetic make-up) breeding population for any one species to survive in a given range.

Predator. An animal that obtains its food primarily by killing and consuming other animals.

Primary (or natural) forest. A forest largely undisturbed by human activities.

Primary productivity. The transformation of chemical or solar energy to biomass. Most primary production occurs through photosynthesis, whereby green plants convert solar energy, carbon dioxide, and water to glucose and eventually to plant tissue. In addition, some bacteria in the deep sea can convert chemical energy to biomass through chemosynthesis.

Protected Area (PA). An area of land and/or sea especially dedicated to the protection and maintenance of biological diversity, and of natural and associated cultural resources, and managed through legal or other effective means.

Provinciality effect. Increased diversity of species because of geographical isolation.

Recalcitrant seeds. Seeds that cannot survive the reductions in moisture content or lowering of temperature necessary for long-term storage.

Recombinant DNA technology. Techniques involving modifications of an organism by incorporation of DNA fragments from other organisms using molecular biology techniques.

Rehabilitation. The recovery of specific ecosystem services in a degraded ecosystem or habitat.

Restoration. The return of an ecosystem or habitat to its original community structure, natural complement of species, and natural functions.

Riparian. Related to, living, or located on the bank of a natural watercourse, usually a river, sometimes a lake or tidewater.

Seedbank. A facility designed for the ex situ conservation of individual plant varieties through seed preservation and storage.

Selection. Natural selection is the differential contribution of offspring to the next generation by genetic types belonging to the same populations. Artificial selection is the intentional manipulation by man of the fitness of individuals in a population to produce a desired evolutionary response.

Serological testing. Immunologic testing of blood serum for the presence of infectious foreign disease agents.

Somaclonal variations. Structural, physiological, or biochemical changes in a tissue, organ, or plant that arise during the process of in vitro culture.

Species. A group of organisms capable of interbreeding freely with each other but not with members of other species.

Species diversity. The number and variety of species found in a given area in a region.

Species richness. Is the number of species within a specified region or locality.

Spectroscopy. Any of several methods of chemical analysis that identify or classify compounds based on examination of their spectral properties.

Stochastic. Models, processes, or procedures that are based on elements of chance or probability.

Subspecies. A distinct form or race of a species.

Succession. The more or less predictable changes in the composition of communities following a natural or human disturbance.

Sustainable development. Development that meets the needs and aspirations of the current generation without compromising the ability to meet those of future generations.

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

Systematics. The study of the historical evolutionary and genetic relationships among organisms and of their phenotypic similarities and differences.

Taxon (pl. taxa). The named classification unit (eg Homo sapiens, Hominidae, or Mammalia) to which individuals, or sets of species, are assigned. Higher taxa are those above the species level.

Taxonomy. Is the classification of animals and plants based upon natural relationships.

Threatened species. A United States technical classification referring to a species that is likely to become endangered within the foreseeable future, throughout all or a significant portion of its range.

Tissue culture. A technique in which portions of a plant or animal are grown on an artificial culture in an organised (eg as plantlets) or unorganised (eg as callus) state.

Trophic level. Position in the food chain, determined by the number of energy-transfer steps to that level.

Variety. See 'Cultivar'.

Wild relative. Plant species that are taxonomically related to crop species and serve as potential sources for genes in breeding of new varieties of those crops.

Wild species. Organisms, captive or living in the wild, that have not been bred to alter them from their native state.

Wildlife. Living, non-domesticated animals.

7.2.2 Information Management Terms

Application. Any special purpose software fulfilling a specific function on the desktop. Applications can be general-purpose (eg a word processor) or custom-built to meet a specific requirement.

(Database) Application. A collection of tools (eg data entry screens, reports) which facilitate the operation of a database.

American Standard Code for Information Interchange (ASCII). A standard character set that assigns a numeric code to each letter, number, and selected control characters.

Attribute. Properties of an entity which are measured to produce data (eg 'designation' is an attribute of the 'Protected Areas' entity).

Benchmark. A numerical value that gives a measure of the performance of a computer product in a specific test.

Best Practice Technology (BPT). The compromise whereby industrial premises are allowed to emit higher than normally acceptable pollution levels due to exceptional circumstances. these circumstances include the use of equipment which in itself is not life-expired, they are using in effect the best practicable means available to them.

Bulletin board. Also known as a newsgroup, is an 'area' on a WAN where text messages can be posted by an author, so that they are available to be read by anyone accessing the bulletin board.

CD-ROM (Compact Disc-Read Only Memory). A relatively new technology that uses laser-read discs with their high data compression to store very large amounts of data. Data can only be read from the disk, it cannot be altered or re-written.

Central Processing Unit (CPU). The microchip that is the 'computer within the computer', it logically coordinates the operations of all the other components of the computer.

Client-server. A computer architecture that is a hybrid of the traditional stand-alone and network options with computing tasks shared between the server and the user's workstations.

Computer Aided Design (CAD). Software used for designing in general. It facilitates geometrical drawing on the computer.

Computer Aided Software Engineering (CASE). Software used for designing and developing information systems and databases.

Data. Facts that result from measurements or observations.

Database. A logically structured and consistent set of data that can be used for analysis.

Database Management System (DBMS). Software which stores, maintains, and retrieves data. May also offer a wide range of additional features for data analysis and management.

Data Definition (or Description) Language (DDL). A programming language used to describe the structure and content of data files and the relationship between them (often referred to as schemas). A data description language is included as one component of many database management systems.

Data dictionary. A repository of information about the definition, structure, and use of data.

Data flow model. A representational tool showing how information flows in an organisation or process. Special symbols depict different kinds of flow.

Data model. A representational tool showing the structure and inter-relationships between data entities.

Dataset. A collection of data and accompanying documentation which relate to a specific theme (usually consisting of one or more computer readable files on the same system).

DBF format. The data file format originally used by the dBASE product and now the most common PC DBMS format.

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

Directory Interchange Format (DIF). A data structure originally defined by NASA used to exchange directory - level information about data sets among information systems.

Dynamic Data Exchange (DDE). A mechanism of 'live link' which enables items of information in separate application programs to be inter-connected.

Electronic mail (email). A network (including Internet) resource allowing messages and files to be sent and received between computers.

Entity. Items of interest (concrete of abstract) whose attributes (properties) are being measured.

Entity-Relationship (E-R) diagram. A respresentational tool showing the relationships between entities in an information system.

Field. In the context of databases, a field is a vertical column in a database table.

File Transfer Protocol (FTP). An Internet resource allowing exchange of files between remote computers.

Flatfile. A matrix of columns (fields) of data, where each row represents one record. Equivalent to the term 'Table' or 'Relation' in a relational database.

Flat-file database. The simplest type of database that allows the user to work with only one table of data ('flat-file') at a time.

Geographic Information System (GIS). An information system that stores and manipulates data which is referenced to locations on the earth's surface, such as digital maps and sample locations.

Geo-referenced data. Data which is connected to a specific location on the Earth's surface.

Global Positioning System (GPS). A data capture tool allowing mobile receivers to determine their position anywhere on the Earth's surface in latitude and longitude coordinates to an accuracy of fractions of a second of arc (1 second of arc latitude is approximately 30 metres).

Graphical User Interface (GUI). Computer software that is controlled by the user by the selection of options and symbols from a pictorial presentation on the computer screen (Microsoft's Windows is the most frequently seen example). The contrasting approach is a 'command line' interface.

Hard copy. Data or information that has been printed out from a computer onto paper.

Hardware. The physical components of a computer system such as the computers, disk drives and the screen.

Hyperlink. Hyperlinks are connections that have been programmed into a 'hypertext' document. A reader browsing a hypertext document can select a hyperlink symbol to be presented with additional text on the subject of interest.

IBM compatible. Describes equipment, ranging from personal computers to large mainframes, that can run operating or applications software written for equivalent IBM computers without alteration.

Index. A direct access method to data in a database. An index has a key value and a pointer to the row of the table that contains data with the key.

Information. Data which have been interpreted to facilitate understanding.

Information system. A structured set of people, processes, data and tools, for converting data into information.

Interface. The way that users communicate with a computer system.

Internet. The most widely used international communications computer network.

Listserver. An Internet facility similar in concept to a bulletin board. The main difference is that each time a message is posted by an author to a listserver, it is posted out by electronic mail to all the subscribers of that listserver.

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

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

Mainframe. A multi-user computer designed to meet the needs of a large organisation; a mainframe has a greater capacity than that of a minicomputer or a microcomputer.

Menu. A list of options graphically presented for selection to the software application user.

Metadata. Data about data, for instance its location, source, content, or other specifics. Also co-data.

Metadatabase. A database which is designed to manage metadata.

Modem. A piece of equipment used to link digital devices such as computers to an analogue telephone line. The term is a contraction of modulator-demodulator.

Multimedia. Integration of many forms of data in an application, including text, sound, graphics, and video.

Multitasking. A computing environment that allows several software packages to be run concurrently.

Network. A collection of computers that can communicate with each other.

Normalisation. In the context of databases, the process of organising data into a structure of one or more tables, where each column has a specific unambiguous meaning. Normalisation is necessary to achieve the optimum structure for a relational database.

Object Linking and Embedding (OLE). A feature to transfer and share information between different software applications. For example, whilst within a word-processing document, a spreadsheet table can be directly worked upon using OLE.

Object Oriented (OO). A way of looking at processing problems and their solutions in terms of 'objects'. An object has a recognisable identity which includes information on its 'behaviour' and function. In contrast with conventional software where program and data are separated, the object includes both the data and the procedures and functions that operate on it. Objects cooperate by sending messages to one another.

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

Operating system. Controls access to all the resources of the computer and supervises the running of other programs. Examples of operating systems are MS-DOS, Windows and UNIX.

Optical Character Recognition (OCR). Technique for rapid capture of text into a computer. First the text is scanned, then the image of each character in the text is analysed and converted into the computer code. Characters that cannot be matched may be displayed on a screen for an operator to enter manually.

Personal Computer (PC). Otherwise known as a microcomputer, is a single-user computer with a central processing unit based on a microprocessor chip.

Physical database. The actual physical structure of databases as implemented for a particular hardware or software configuration and database system.

Pixel. Abbreviation for picture element, meaning the smallest, discrete elements that are used to create an image on a visual display unit.

Polygon Attribute Table (PAT). The database table associated with a spatial dataset holding details (attributes) of the geographic objects.

Process. An activity, function or procedure applied to a resource (eg an arithmetic procedure applied to data, or a critical step in a business operation).

Process model. A representational tool consisting of language and diagramming standards representing the inter-relationships between a group of related processes.

Prototyping. A system development methodology which quickly develops a partial or preliminary version to determine its feasibility and user evaluation. Prototypes can then be refined into delivered applications.

Public domain. Intellectual property available to people without paying a fee. Most computer software developed at universities is in the public domain.

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

Random Access Memory (RAM). Dynamic memory provided by the computer's RAM microchips, sometimes known as central memory or core.

Raster graphics. Definition of an image to be produced on a computer screen is stored on a 'pixel-by-pixel' basis.

Record. A collection of data about a specific case or subject. In the context of databases a record is a horizontal row in a database table.

Relational database. A database consisting of two or more tables related via common fields.

Relational Database Management System (RDBMS). Advanced DBMS software which allows the storage of multiple, related files.

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

Server. Any program or computer that provides a service to other programs or users. A network server, for example, provides dedicated hardware and software for the purpose of giving terminals or computers access to a network.

Software. The programs that are run on a computer.

Spatial data. Data which contains reference to a location (which may be a specific location on the Earth's surface, or relative to an arbitrary point).

Spreadsheet. A software program that allows users to establish relationships between rows and columns of data in a tabular format.

Structured design. A methodology for the design of information systems that breaks the program down into a series of modules with carefully specified interfaces between the modules.

Structured Query Language (SQL). ANSI standard data manipulation language used in most relational database systems.

Table. An physical entity in a relational database, in which data are laid out in rows and columns.

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

Vector graphics. Definition of an object's image to be produced on a computer screen is stored by defining its geometry as a series of connected points - to be contrasted with raster graphics.

Wide Area Information Server (WAIS). A system designed for retrieving information from networks. It is a searching facility dependent on matching requests with a specific request.

Wide Area Network (WAN). A computer network where the constituent systems may be widely dispersed geographically and links are formed by the use of telephones, radio, satellite, etc.

Workstation. Powerful desktop computer equipped with a high-resolution display and designed for technical applications. Groups of these workstations are normally linked to a shared computer which holds common information.

World Wide Web (WWW). Popular Internet resource based on the exchange of information via a graphical, hypertext, interface.

Universal Resource Locator (URL). Address describing the location of information sources on the Internet global communications network.

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