THE BIODIVERSITY INFORMATION CLEARING HOUSE

CONCEPT AND CHALLENGES

World Conservation Monitoring Centre
219 Huntingdon Rd, Cambridge, CB3 0DL

Supported By The UK Department of Trade and Industry.

TABLE OF CONTENTS

- 1. INTRODUCTION
  - 1.1 THE NEED FOR A CLEARING HOUSE FOR THE CONVENTION
  - 1.2 THE NATURE OF THE ISSUE
  - 1.3 WHAT IS A CLEARING HOUSE?
- 2. THE NATURE OF BIODIVERSITY INFORMATION
  - 2.1 SCOPE
  - 2.2 FORMS OF BIODIVERSITY INFORMATION
  - 2.3 SUMMARY
- 3. THE STATE of EXISTING INFORMATION and RETRIEVAL SERVICES
  - 3.1 OVERVIEW
  - 3.2 SOME EXAMPLES OF EXISTING SERVICES
- 4. EXPERIENCE WITH BIODIVERSITY INFORMATION SERVICES AT WCMC
  - 4.1 INTRODUCTION
  - 4.2 DATA HOLDINGS
  - 4.3 BIODIVERSITY INFORMATION MANAGEMENT EXPERIENCE
  - 4.4 SUMMARY
- 5. CHALLENGES in the DESIGN and OPERATION of RETRIEVAL and REFERRAL SYSTEMS
  - 5.1 TECHNOLOGICAL CHALLENGES
  - 5.2 INSTITUTIONAL AND MANAGERIAL CHALLENGES
- 6. THE WAY FORWARD: A COOPERATIVE CLEARING HOUSE NETWORK
  - 6.1 THE PRINCIPLES
  - 6.2 THE CONCEPT
  - 6.3 PRACTICAL CONSIDERATIONS
- LIST of ACRONYMS
- ACKNOWLEDGEMENTS

1. INTRODUCTION

1.1 THE NEED FOR A CLEARING HOUSE FOR THE CONVENTION

Effective analysis, priority setting and action to protect the World's biodiversity requires the manipulation and integration of a vast range of scientific and socio-economic data into a form suitable for decision making. It is acknowledged that there is a pressing need for those countries (developed and developing) with particular knowledge of one aspect or another to share this information and know-how with others for mutual benefit.

The Convention on Biological Diversity (CBD) in a number of its Articles requires or implies the need for facilities for the management and open exchange of biodiversity information. Articles 7d, 12c, 13b, 15(7) and 16, each identify information management and exchange requirements, and Article 17 explicitly indicates "*access to and transfer of technology among Contracting Parties are essential elements for the attainment of the objectives of this Convention*". Article 18(3), requires the establishment of "*a clearing house mechanism to promote and facilitate technical and scientific cooperation*". Some central focus through which exchange of information can take place is clearly both highly desirable for, and explicitly required by, the Convention.

The CBD identified three main categories of biodiversity information which will be required:

- ecosystems and habitats,

- species and communities, and

- described genomes and genes of social, scientific or economic importance.

To this basic list one must add:

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

- bio-technology, its value and risks;

- local knowledge of traditional uses and values of biological resources;

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

One way of sub-dividing the information requirements under the CBD is in the following eight major categories which have some pragmatic appeal, as they align with the way in which national and international institutions are organised to manage information:

Conservation
: encompassing information on species, habitats, protected areas, biodiversity indicators, wildlife, ...

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

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

Bio-technology
: encompassing a forum for interchange of information on research and application of bio-

technology, ...

Environmental Statistics/Economics

encompassing resource utilisation, value of biodiversity, land use, industrial outputs, natural resource utilisation, trade, economics, ...

Policy

encompassing policy development, modelling, decision support systems and technology, empowerment and public consultation techniques, ...

Human Factors

encompassing population, human health, social conditions, and their relationships to biodiversity, ...

Environmental Law

encompassing environmental legislation, conventions, protocols, regulation, standards, ...

The Report of the Open-ended Intergovernmental Meeting of Scientific Experts on Biological Diversity - Second Session (UNEP/CBD/IC/2/11) provides in the Annexes further indication of the scope of the information and technology of biodiversity. For instance, Annex II, relevant to the identification, characterisation and monitoring of ecosystems, species, and genetic resources, lists 6 major categories of technology (and related information), further subdivided into a large number of classes from "*biogeography, through ecosystem function and traditional knowledge through to abundance, distribution and range (of species) and biotechnology*".

With limited resources available, the setting of priorities for the types of information to be collected and exchanged is critical. These will vary according to needs and requirements, and must be subject to a thorough assessment of the costs and benefits involved. Before developing any new information management and exchange operation, it is essential to review what information already exists, who is collecting it and how, who is managing it and how, and so on to avoid duplication and facilitate cooperation.

In the past, there has been a tendency to assume that it is best to collect and manage all information in one central location so as to ensure its adequate integration and application. For various reasons this may not be the best solution. What is more important is the application of appropriate standards to data collection and management, custodianship by the appropriately qualified centre of excellence, good communications between information collectors and managers, and coordination of information collection and management effort.

1.2 THE NATURE OF THE ISSUE

National government policy, research, and resource management organisations, NGOs, researchers and international environmental agencies have been collecting information on various aspects of biodiversity for many years, and in the last few decades, with satellite technology, data volumes are exploding. In parallel, both the technology of data management and exploitation of biological resources (**bio-technology**) have advanced enormously. As a result, a great deal of information currently exists - including detailed observational data, traditional experience and know-how, technology, and technical know-how.

Access and effective use of this information is hampered by:

- lack of shared knowledge of who has what information, technology or expertise, as it is held in a broad range of separate sectoral and regional institutions;

- lack of consistency in observational methodologies, classification standards, quality assurance methods and analytical models, so that inter-sectoral and cross-sectoral integration and comparison are very difficult; and

- lack of equity of access due to varying levels of information technology infrastructure, and "*information buying power*".

1.3 WHAT IS A CLEARING HOUSE?

In its original meaning (19th C), a **Clearing House** was an establishment where financial institutions adjusted claims for cheques and bills and settled mutual accounts with each other. The term has extended in meaning to refer to mechanisms of sharing information (not only financial) through an independent intermediary for mutual benefit of the concerned parties. Thus a clearing house to "*facilitate technical and scientific cooperation*" (in biodiversity) is a facility for the deposit and open exchange of information for mutual benefit on the state of biological resources, and related data collection, research, and technology. The key characteristics are :

- **independence** (of the clearing house), that is the operators of the clearing house should have no vested interest in controlling or subverting the flow of information; and

- **mutual benefit**, which usually means, in practical terms, that no detailed accounting is made of transactions; equality of contribution and benefit is assumed to occur over a long period.

The modern clearing house is usually aided by computer networking and information management technology, to allow for the exchanges to take place rapidly and efficiently. One implication is that there is usually a store of information which has been contributed to the clearing house. With network technology this store may be conceptual, that is, the information continues to physically reside with the **contributing** agencies or subsidiary clearing houses, who continue to exercise expert custodianship.

The purpose of a biodiversity clearing house is to alleviate the challenges identified in Section 1.2 in an efficient manner - for instance, providers of information or technology should only need to provide it once, and not have to respond to repetitious requests, and information seekers should have a single window source. In this way the goal of "*facilitating technical and scientific cooperation*" would be achieved.

A background paper to the "*Informal Consultation on how to establish a Clearing House Mechanism*" hosted by the interim secretariat to the CBD, Geneva, August 26-27 1994, identified four models for a clearing house mechanism:

- **information** retrieval and referral services;

- demand based **broker** services where, upon request, an intermediary assists service users with particular needs by providing advice, finding technical or financial partners or both;

- **technology** development services which distribute technology to partners for local adaptation and feed back; and

- **policy** development services which provide assistance upon request.

This background paper concentrates on the first of these - the **information service** clearing house, although in practice any clearing house usually operates to some extent in more than one of these modalities. It should be noted however that there are more proactive elements to brokerage, technology development and policy development services which greatly add to the complexity of operation and cost of infrastructure, than the more passive information retrieval and referral style of clearing house.

A very important aspect of any form of clearing house is the need for **meta-information** (or meta-data), that is information which describes information (eg; its nature, quality, source, formats, and **accessibility**). An information service which mainly holds meta-data which **refers** the query to the source of the information is often called a referral system. Bibliographic and abstracting services are of this nature, and INFOTERRA is a well-known example of a referral system in the environmental field.

An **information retrieval service** allows the user to obtain the desired data directly without (apparent) referral to another service or agency. In practice, many services may operate in both modes - for instance providing direct retrieval of smaller datasets or documents, while referring the user to the source agency for large databases or non-digital records, such as maps and books.

The key to a successful information retrieval and referral service is good meta-data, organised in a systematic and controlled manner. The meta-data must in turn be supported by controlled vocabulary, thesauri, data dictionaries and the like, which aid in meeting the user query specification by providing information on coded fields, keywords, synonyms (multilingual equivalents, if possible), broader and narrower related terms etc.

Thus, the basic components of an information service clearing house are:

- meta-database
- thesaurus of terms
- data dictionary
- information registry and quality control system
- referral system
- information retrieval and management system
- user query interface
- electronic network connections.

Tools to update and maintain all of the above also need to be in place.

2. THE NATURE OF BIODIVERSITY INFORMATION

2.1 SCOPE

As noted in Section 1.1, and implied in the documents relating to the CBD, the range of potential information types varies widely, with biodiversity factors placed in the context of a range of other types of information necessary for the management and understanding of biodiversity. The following paragraphs, adapted from the United Nations Environment Programme (UNEP) Guidelines for Country Studies (1993), identify some of the types of information which might be important, particularly for the initial assessment and strategy development.

Biological

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

Physical

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

Socio-economic

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

Costs and benefits

In order for management of biodiversity to be efficient it is necessary to know the true value of biodiversity, and the costs and benefits of management options. This needs to cover questions such as the costs of managing protected area systems, the level of income derived from visitation, and the value of indirect benefits derived from protection of the watershed. Some of these values require methods of assessing the true benefits of biological diversity which are only just being developed, and wider dissemination of information on the methodologies will be required.

Threats

Identifying and monitoring both potential and actual threats to biological diversity is likely to be an important component of any data collection and management programme oriented towards improved management of biological resources. However, information collection programmes may need to go beyond the obvious physical causes and their effects to also identify and record the underlying causes, including the responses to human activities (which links threats to socio-economic factors).

Management

Conservation of biological diversity is about effective and sustainable resource management. To assess that management, information will not only be required on the biodiversity, its status and its distribution, but also on current and past management activities, especially on the **use** of biodiversity resources. For example, information is likely to be required on a range of characteristics of protectedareas and on effective management methods and technologies in a range of habitats, both protected and unprotected.

Sources and contacts

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

Interrelationships

The above paragraphs, although brief, begin to indicate the extent of the interrelationships between the information that might be required in order to study and manage biodiversity more effectively. It is essential that these interrelationships are kept clearly in mind when planning both information collection and management strategies, and information on methods of assessing and forecasting effects of interrelationships must form an important component of any information sharing mechanism.

All this points to a vast scope for the information to be collected and exchanged, and clearly some selection must be made. The following is a suggested list of the most relevant categories:

- information which provides a practical baseline for monitoring effectiveness of action

- key information identified by biodiversity managers as being important for decision making

- socio-economic value of the local and national biodiversity, and of protected areas

- functions and benefits of biodiversity, particularly service functions of ecosystems and protected areas

- policy and conservation programmes, and the legislative framework and other institution-related matters

- technology useful in monitoring, assessing and improving the sustainable utilisation of biological resources

- genetic resources, including medicinal plants, landraces and wild ancestors of domestic breeds and cultivars

- species that could serve as indicators of ecosystem health and *flagship* species habitats, the conservation of which protect the diversity of other species and habitats

- alien or exotic species, the spread of which could threaten indigenous biological diversity

- threats to biodiversity and biodiversity known to be threatened

- species and habitat distribution and its change with time

2.2 FORMS OF BIODIVERSITY INFORMATION

The information required occurs in a range of forms and media, including:

Numeric

Numeric data is derived directly from many types of survey ranging from counts of species in particular locations, 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 recording machines, or derived from analysis of remote sensed data.

Because of its nature, numeric data lends itself to computer aided manipulation and 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. It is 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 quite strictly, for instance into database tables, and to exercise quality checking on data entry against norms, and various types of statistical analysis may be applied.

Categoric

A very common form of biological information is classified or coded non-numeric data, and might include indications of factors such as soil type, land cover, forest type, species, protected area designation, and so on. The data is structured, usually through a thesaurus or data dictionary, and can be checked for allowed values, and the like. Statistical analysis is not often appropriate, but spatial distribution is often important.

Text

Text is an extremely common form of biodiversity information, including descriptions of protected areas, descriptions of species, descriptions of threats, ecosystem status reports, "*State of the Environment*" reports, legislation, regulation, strategies and plans. By its nature, it is much less structured, is often subjective, and difficult to appropriately search and retrieve unless careful controls are placed on its compilation. However, with recent developments in text manipulation using computers, such as the use of *hypertext links*, the value of text description has increased considerably.

Text, when attached to a database of numeric and categoric information, can provide valuable extra description on the quality and sources of the data. For example, it can be used to check more thoroughly the identity of a species or expand upon an ecozone classification once a database has been used to identify it.

Spatial data

Biodiversity information is intrinsically spatial in nature, including the locations of sample points, mappable phenomena such as climate, topography, habitat, species range, point located and distributed threats, vegetation and land use. This information is often contained on paper maps, in digital remotely sensed imagery or in computer-based geographic information systems.

Other Media

Biodiversity information also includes non-digital images (photographs, drawings) of landscapes, specimens, and technology (instruments, methodological flow diagrams and the like). Future consideration should also be given to moving images (such as video tapes recording wildlife behaviour) and sound recordings.

2.3 SUMMARY

The information that a clearing house must deal with has the characteristics of being:

- multi-sectoral - including many sectors of **environmental** and natural resource data, as well as socio-economic and policy information;
- spatial - often over very large areas of the globe;
- diversified - containing statistical data, quantitative scientific measurement, numerical and categoric classified data, and descriptive/narrative information; and
- very voluminous - millions of documents, terrabytes of data.

The technical challenges to an effective clearing house mechanism include:

- effective management of very large spatial databases;
- abstraction and summarisation in a manner which facilitates decision making;
- harmonisation of information from differing measurement regimes;
- integration of data from different environmental and socio-economic sectors;
- maintaining quality control and metadata on the quality and origins of the information; and
- integration of textual information with conventional databases.

3. THE STATE of EXISTING INFORMATION and RETRIEVAL SERVICES

3.1 OVERVIEW

Bibliographic referral systems covering a range of scientific topics including environment and biodiversity have been well established since the 1960s, and UNEP's INFOTERRA is a good example. Recently, a series of five directories and surveys from the UNEP Office of Harmonisation of Environmental Measurement (HEM) have identified over 300 significant institutions world-wide which collect or manage environmental information. A recent survey conducted on behalf of the European Environment Agency (EEA) located 180 institutions in the UK alone which dealt with environmental data.

Coordinated by the International Council of Scientific Unions (ICSU), the World Data Centers (WDC) have acted as a clearing house for geophysical, oceanographic and other scientific data since the International Geophysical Year. New initiatives such as the Consortium for International Earth Science Information Network (CIESIN) have broadened the scope, particularly for the electronic transfer of digital data.

The list of established data centres and clearing houses is quite long, although there are cautionary notes - many are quite specialised (such as the Environmental Law Centre) and those dealing with the more manageable, and quantitative information characteristic of oceanographic and atmospheric studies are more advanced in standards and coordination than those in terrestrial ecology. The move towards a CBD clearing house mechanism is needed to bring this sector more in line with other scientific sectors.

In spite of the many efforts on information sharing currently under way, the task of biodiversity information sharing (alone, without adding the complexity of the brokerage of technology cooperation) is not to be underestimated because of the huge scope and current lack of agreement on information requirements, standards and harmonisation principles. This is all the more reason why it is important to build on the existing experience, infrastructure and institutions, some examples of which are given in the next section.

3.2 SOME EXAMPLES OF EXISTING SERVICES

The clearing house mechanism may wish to consider existing organisations within the UN system, as well as NGOs. The UN organisations such as Food and Agriculture Organisation (FAO), United Nations Educational, Scientific and Cultural Organisation (UNESCO), United Nations Industrial Development Organisation (UNIDO), UNEP and World Health Organisation (WHO) have become established as centres of excellence for developing and managing certain information resources relevant to biodiversity. For example, FAO have been active in developing monitoring and reporting standards for land use, whilst the WHO have custodianship of important databases on demography and health. Similarly there are NGOs including the International Plant Genetics Resources Institute (IPGRI), the World Federation for Culture Collections (WFCC), the World Conservation Monitoring

Centre (WCMC), the Environmental Law Centre (ELC) which manage and distribute important information for the biodiversity community.

Brief descriptions of some of the more relevant services of a general nature are listed below.

CIESIN

The Consortium for International Earth Science Information Network (CIESIN) is an international initiative that serves scientific, policy-making, educational, and public access data and information needs. CIESIN archives and disseminates data relevant to understanding human interactions with the environment. CIESIN services support interdisciplinary research and applications in key areas including global environmental change and sustainable development. CIESIN has organised and maintains access to the Information Cooperative, a distributed global network of data centres and other institutions.

The CIESIN Catalog Service provides tools for users to search and order data ie; an **information retrieval** service. The Catalog Service is accessible via the Internet as well as modem dial-in access. Catalog Servers communicate with each other and are designed to function in a distributed environment - CIESIN's Information Cooperative. The Information Cooperative will enable seamless access to Servers provided by resource centres that archive data ranging from biodiversity to socio-economic research.

CIESIN is collaborating on infrastructure and data development programs with the National Aeronautics and Space Administration (NASA), the U.S. Environmental Protection Agency (EPA), the U.S. Department of Agriculture (USDA), and other US and international agencies, including the United Nations Development Program (UNDP) Sustainable Development Network (SDN).

GLIS

The Global Land Information System (GLIS) is an interactive computer system developed by the U.S. Geological Survey (USGS) for scientists seeking information and access to data pertaining to the Earth's land surface that can be used in continental and global scale Earth science and global change studies. GLIS can be characterised as a metadata system containing both descriptive information and query functions that allow scientists to assess the potential utility of data sets, determine their availability, and place on-line requests for related data products. Both textual and graphical user interfaces are provided, and scientists can access GLIS through either wide-area network or dial-up communications interfaces.

INFOTERRA

The Stockholm Conference on the Human Environment, convened by the United Nations in 1972 to consider the condition of the environment, called for an international mechanism for the exchange of environmental information. The result was The International Referral System (IRS), later renamed The Global Environmental Information Exchange Network or INFOTERRA, which was established in 1975. The main direction given to INFOTERRA was to develop a mechanism to "*facilitate the exchange of environmental information within and among nations*".

INFOTERRA began its operations in 1977 with a dozen partner countries. It was, from the start, designed as a decentralised information system operating through a worldwide network of national environmental institutions designated and supported by their governments as national focal points and coordinated by UNEP Headquarters in Nairobi. Today, this linking structure consists of 165 national focal points, 11 regional service centres, and 34 special sectoral sources. In the early years INFOTERRA operated only as a referral system. However, following the recommendations of an independent assessment of the system done in 1981, INFOTERRA evolved and expanded its services to include substantive information and document delivery.

The INFOTERRA national focal points are usually situated in the information and documentation sections of environment ministries, and national environmental protection agencies which are often also the focal points for national information networks. They act as the primary access points through which queries from users are channelled to INFOTERRA sources and through which users receive their replies.

Each national focal point compiles a "*Who's Who*" of environmental expertise in their country, and selects the best sources for inclusion in INFOTERRA's main publication the *International Directory of Sources*. These are constantly monitored and updated. The International Directory of Sources exists in both printed form and as a database. For this reason the International Directory is often referred to as the INFOTERRA Database. It is a referral system which helps to access more than 7,000 sources of information on over 1,000 environmental subjects. The sources are located in government ministries and documentation centres, research institutes, universities, non-governmental and international organisations, United Nations agencies and private consultancies. Regional service centres have been set up, within key national focal points, to act as centres for regional co-operation for the exchange of information and for the development of sub-networks to facilitate this exchange.

In addition to the International Directory, INFOTERRA periodically publishes specialised directories and sourcebooks like the Thesaurus of Environmental Terms, Operations Manual, quarterly Bulletins, and the technical Exchange of Environmental Experiences Series. The Thesaurus, in particular, is set to become an increasingly important publication as the movement to exchange environmental information gathers momentum. Organisations, such as CIESIN and EEA, who are in the latter stages of implementing their own systems to catalog environmental data held by institutions, are finding that consistency in the use of terminology is perhaps the main obstacle to information exchange. CIESIN have their own short lists of recognised terminology that they inherited from NASA (the organisation that defined the Directory Interchange Format). But these are primarily of relevance to space science and for environmental metadata they are suggesting that the INFOTERRA Terms are used.

The direction for INFOTERRA over the next decade has been further defined by the United Nations Conference on Environment and Development (UNCED) held in Rio de Janeiro in 1992. Chapter 40 of UNCED's global plan of action, Agenda 21, addresses the importance of information for decision making. Part A of that chapter is mostly concerned with data collection, but Part B entitled *Improving information availability* specifically cites expansion of the INFOTERRA system to a world mandate.

EEA

The establishment of the European Environment Agency (EEA) was agreed upon at a March 1990 meeting of the European Ministers' Environment Council in Bruxelles. EEA was conceived as a smaller coordinating unit of a large decentralised network.

The EEA's main aim is to assist in harmonising data and to distribute to European Union institutions, Member States and the public, information about the environmental data which exist in Member States and which are relevant to the Union. The Agency will achieve this through close cooperation with the Member States and relevant international institutions, who actually collect and hold the data.

To perform its tasks, the Agency must have at its direct disposal a limited nucleus of European environmental data and a wide knowledge of sources to other relevant environmental data in Member States and international organisations. It must also create tools and guidelines to assist the harmonisation. The Agency needs to know who holds data, about what and how the data are accessible.

Collecting and distributing knowledge about relevant sources of environmental data and information is to be done through a Catalogue of Datasources (CDS), based on collaboration mainly between the Agency, the National Focal Points of the Member States and a limited number of international organisations. For the EEA-CDS system, a database is currently being built and served by the Agency Task Force staff.

WORLD DATA CENTERS

The WDCs were originally established to store information from ICSU's 1957 International Geophysical Year. An ICSU-WDC Panel is responsible for coordinating the activities of the individual data centres. Currently, 27 WDCs are active collecting, archiving, and disseminating data which encompass most facets of the global environment. WDCs are generally co-located with national data centres and are funded by the respective nation. The U.S.A. (designated WDC-A) sponsors nine centres, while Russia (WDC-B) operates two and 16 other WDCs (WDC-C) are located in other nations, including China, Japan, Switzerland and the U.K.

Data are acquired from various sources and managed according to internationally recommended procedures. For example, the WDC for Greenhouse Gases (WDCGG) in Japan, collects its data from world-wide sources pertaining, in particular, to atmospheric concentrations of CO_2, CH_4, CFCs, and NO_x. Generally information can be obtained from these centres for a small cost.

Given advancing computer technology, the ICSU WDC Panel is developing a revised workplan for the Centres. In particular, emphasis is being placed on the improved electronic exchange of datasets. Some of the larger data sets are available on CD-ROM.

GENIE

The Global Environmental Network Information Exchange (GENIE) is still under development and will be a vital part of the UK contribution to international science programmes including projects within the International Geosphere-Biosphere Programme (IGBP), the World Climate Research Programme (WCRP), and the Human Dimensions of Global Environmental Change Programme (HDGECP).

GENIE is based on the design of a metadata retrieval and management system produced by the Midlands Regional Research Laboratory (MRRL), and is intended to be a fully-distributed system designed to allow those with knowledge of data holdings to make their knowledge available to other researchers. Although no single centre will hold all the information available in the system as a whole, GENIE will assist users to answer a variety of queries.

The GENIE project will provide a user-sympathetic system for locating and accessing relevant information on Global Environmental Change. The software is designed to run on a range of hardware platforms in order to provide information on data availability, location, currency and quality, as well as offering links to other international data directories. The system will have a flexible and intuitive user interface that will allow both enquirers and data suppliers to interact using their own terminology.

WCMC

The World Conservation Monitoring Centre, jointly supported by UNEP, the World Wide Fund for Nature (WWF) and the World Conservation Union (IUCN) has been managing information in the field of biodiversity conservation for more than 12 years. The Centre's experience as a clearing house is described in Section 4.

The following are some examples of some more specialised information services.

IOC

The Intergovernmental Oceanographic Commission (IOC) was established in 1960, built on the foundation laid by UNESCO's International Advisory Committee on Marine Sciences.

The IOC is the only body within the United Nations charged with basic oceanographic research. Its functions are to develop, recommend and co-ordinate international programmes for scientific investigation of the oceans and to provide related ocean services to Member States.

The IOC promotes marine scientific investigations in the fields of ocean dynamics, circulation, climate and living resources through specialised research and monitoring programmes.

The IOC/UNEP programme on Global Investigation of Pollution in the Marine Environment (GIPME) concentrates on studies of pollution distribution and impacts as well as inputs of pollutants to coastal zones and open ocean areas. The scientific basis for this programme and the associated activities are developed through three Groups of Experts: GEMSI (Methods, Standards and Intercalibration), GEEP (Effects of Pollutants) and GESREM (Standards and Reference Materials).

ELC

The Environmental Law Centre (ELC), the *legal arm* of the IUCN Secretariat, monitors and maintains databases on legal trends and developments in the environmental field, including international agreements, binding instruments of international organisations, national legislation and legal literature. It also develops specific databases (eg; on species protection); contributes to the work of other organisations working in this field; supports activities of other IUCN components; and develops and carries out specific legal activities (eg; drafting international treaties).

The Environmental Law Information System (ELIS) is developed, maintained and operated by the Law Centre staff and is geared to give information to people throughout the world. ELIS is composed of the following four main databases:

- national legislation (approx. 37,000 records);

- international treaties (approx. 800 records);

- supranational instruments (approx. 350 records);

- law and policy literature (approx. 39,000 documents).

IRPTC

In 1976, UNEP established the International Register on Potentially Toxic Chemicals (IRPTC) in Geneva. As the IRPTC is set up as a world-wide network, its co-operation with other national and international institutions is extensive. The UNEP Governing Council gave the IRPTC four main objectives, and a fifth was added in 1989. These objectives are:

- to facilitate access to existing data on the production, distribution, release and disposal of chemicals and their effects on humans and their environment, thereby contributing to a more efficient use of national and international resources available for the evaluation of the effects of the chemicals and their control;
- on the basis of information in the Register, to identify the important gaps in existing

knowledge on the effects of chemicals, and call attention to the need for research to fill those gaps;

- to identify, or help identify, potential hazards from chemicals and wastes, and to improve awareness of such hazards;
- to provide information about national, regional and global policies, regulatory measures and standards and recommendations for the control of potentially toxic chemicals; and
- to facilitate the implementation of policies necessary for the exchange of information on chemicals in international trade.

IRPTC is functional through a network of national and international organisations, industries and external contractors, and national correspondents. To date, IRPTC has correspondents in approximately 110 countries. Within IRPTC, the UNEP Governing Council established two information services. The first is the IRPTC Legal File, and the second is the IRPTC Query-Response Service.

The IRPTC Legal File was established in the late 1970s and published its first *Legal Data Profiles for Selected Chemicals* in 1980. This File was established primarily to manage special files on waste management and disposal, chemicals currently being tested for toxic effects, and national chemical restrictions.

The Query-Response Service was established to answer questions from national governments, environmental authorities, industry and individuals world-wide.

The files are available through on-line networks from both the Environmental Chemicals Data and Information Network of the EC (ECDIN) through EURONET, TELEPAK, and TYMNET; and from the Department of National Health and Welfare (Canada) through the National Telecommunications Network (INET 2000). Information can also be accessed, free of charge, directly from IRPTC through the Query-Response Service. Different forms of information dissemination are also being examined, including storage on CD ROM.

WFCC

The World Federation for Culture Collections (WFCC) is a 200-member federation of the International Union of Microbiological Societies. Microbiologists in 55 countries working in research, education, and industry are encouraged to cooperate in the study of procedures for the isolation, culture, characterisation, conservation, and distribution of microorganisms. The aim is to establish an effective network of individuals and institutions possessing collections of microorganism cultures and cell lines and to facilitate communication between collection owners and users.

The long-term goal of the Federation is to create a global network of information services charged with compiling and disseminating data on cultures; address practical questions such as the impact of postal regulations, quarantine rules, patent laws, and public health concerns on culture distribution.

The WFCC pioneered the development of an international database on culture resources worldwide. The result is the World Data Center for Collections of Microorganisms (WDCM). This data resource is now maintained at RIKEN, Japan and has records of nearly 400 culture collections from 55 countries. The records contain data on the organisation, management, services and scientific interests of the collections. Each of these records is linked to a second record containing the list of speciesheld. The WDC database forms an important information resource for all microbiological activity and also acts as a focus for data activities amongst WFCC members.

Computerised Information Services operated by the Federation also include the Microbial Strain Data Network (World Data Center for Collections of Microorganisms) for communications purposes. Publications include the *World Directory of Collections of Microorganisms* and the *Living Resources for Biotechnology*.

4. EXPERIENCE WITH BIODIVERSITY INFORMATION SERVICES AT WCMC

4.1 INTRODUCTION

The World Conservation Monitoring Centre was founded in 1988 by IUCN, WWF and UNEP. The Centre is an independent non-governmental organisation with 50 staff based in new headquarters in Cambridge, UK. It is governed by a Board comprising two delegates from each founder agency, and an independent Chairman. The Centre's programme is developed with help from an external Programme Advisory Group.

WCMC specialises in providing information on human impacts on global biodiversity, including both ecosystems and species. This task is addressed through a programme of *Data Management and Integration*, an *Information Service,* and *Technical Services*, in close association with its founder agencies, and with a wide range of collaborating organisations and associated experts around the world. The Centre benefits greatly from the founders' international constituencies, in particular IUCN's member governments and voluntary commissions, WWF's links with NGOs and corporations worldwide, and UNEP's links with the UN family and the conservation treaty secretariats. These networks are an essential ingredient of WCMC's work. In return, the Centre's work in gathering, managing and disseminating information reflects the focal points of its founders' programmes.

4.2 DATA HOLDINGS

WCMC has been managing information in the field of biodiversity conservation for more than 12 years, and has built up extensive practical experience in the development of computer databases, and the management of related documentation. WCMC currently collects and manages data for the worlds protected areas, major habitats, species threats and distribution, in traded species (Convention on International Trade in Endangered Species, CITES) and important conservation literature.

The main databases currently available at WCMC are briefly described as follows.

Tropical forest database

In collaboration with IUCN, WCMC has built a database of tropical forest maps, at a working scale of 1:1 million, based on the most recently available sources. The data are held on a computerised geographic information system (GIS) and are available for dissemination in digital or paper format. Some of the information has already been published in two volumes of a three volume series, the *Conservation Atlas of Tropical Forests*. Coverage includes rain, monsoon, montane, swamp and mangrove forests. The flexibility of the GIS enables overlay of the forest cover data with other relevant information, such as national parks coverage or population density, as a means of analysis and assessment. WCMC knows of no other global forest cover database held in digital form, although many partial datasets do exist, and have been used in this compilation.

Protected areas database

WCMC has recently completed a digital database containing maps and statistical data for the world's protected areas system. Data have been published in a four-volume *Directory of the National Parks* *and Protected Areas of the World*. In addition, the Centre has used the GIS to map all those Forest Reserves that are managed for environmental or conservation purposes. The results are being used by FAO as a contribution to the *Forest Resources Assessment 1990 Project*. The overall database now includes data and location maps for over 29,000 protected areas.

Biodiversity Map Library

The digital coverages of protected areas and forests are augmented by a wide range of other thematic GIS data. For example, WCMC has digitised a number of biogeographical classifications, global potential vegetation maps, maps of globally important wetland and certain other ecosystems (particularly coastal habitats such as coral reefs), maps of critical conservation sites noted for their biodiversity but lacking legal protection, and many others. A catalogue of these data is already available. The data are stored, managed and manipulated through a purpose-designed Biodiversity Map Library which is designed to meet users' needs and will soon be available as an on-line service. The Map Library will be a central element in the *International Forest Information Service*.

Plants database

Of the estimated 270,000 kinds of higher (vascular) plants on earth today, over 10% are threatened with extinction. Since the late 1960s, WCMC has gathered and provided information and analyses on plants of conservation concern and is the only organisation to do so on a global scale. Currently, taxonomic, distribution, and conservation information is held on more than 81,000 kinds of plants (species, subspecies and varieties). In WCMC's plant database (the largest plant conservation database in the world), data is recorded on single-country endemics and all other plants of conservation concern at the national level. More recently, data has been gathered on tropical timber species, wild relatives of crop plants, plants of ethnobotanical or pharmaceutical value, plants present in protected areas. Computerised data are stored in a series of relational database files managed by *BG-BASE*, a variable-length field database application built using Advanced Revelation. All data is linked to a computerised bibliography of 17,000 references dealing with plant conservation.

Animals database

WCMC has compiled the *IUCN Red List of Threatened Animals* in collaboration with IUCN Species Survival Commission. All the data on the 6000 animal species currently regarded as threatened globally are held on computer, including their nomenclature, common names, distribution and conservation status. Additional databases list country endemic species. The Centre is active in preparing digital distribution maps for threatened animals and has developed maps for marine turtles and endemic freshwater fishes.

Trade database

The *CITES Trade Database* is the largest database of its kind, holding some 2 million records on trade in wildlife species and their derivative products. It is managed for the Secretariat of the Convention on International Trade in Endangered Species, of which there are currently 122 member states. The information spans from 1975 to the present and information is updated from annual reports submitted by CITES Parties which are then entered either manually or loaded directly into the database. This data includes information on gross or net import/export data for a specified year (s), country, species and/or product. Trade data can be provided by WCMC (with the permission of the CITES Secretariat) in different formats which allow yearly trends to be monitored; comparison between reporting Parties to be made and illegal trade to be identified; and annual reports to be produced for CITES Parties not able to produce their own.

Bibliographic files

WCMC maintains an extensive library that is particularly well stocked with **grey** literature difficult to obtain elsewhere. Records are maintained on a computerised bibliographic system that can be searched and manipulated to meet users' needs.

4.3 BIODIVERSITY INFORMATION MANAGEMENT EXPERIENCE

In the process of developing these databases and services over the last decade, WCMC has gained considerable experience in the application of modern technology to the provision of biodiversity information services. For example, over the past two years WCMC has been collaborating with an international corporation to develop an advanced GIS-based information management system - the Biodiversity Map Library. This system facilitates access to spatially distributed biodiversity information through a user-friendly query system, providing non-expert users much of the power of a computer GIS, without requiring them to be familiar with GIS software and technology. This database offers an advanced take-off point for further projects in this area.

A great deal of the information on species, protected areas and habitats occurs in text form, and considerable expertise has been accumulated in using text searching and indexing tools, and standard data dictionaries to integrate conventional databases with narrative information. An integrated metadatabase has been found to be a powerful and convenient tool for searching and locating information across the disparate spatial, textual and conventional databases.

In order to connect with other information clearing houses, WCMC is increasing its experience and services to customers over the international communications networks. WCMC has been connected to the Internet, which is clearly the most widely used network facility for science in general, for several years. This allows WCMC to access and be accessed electronically by other users of the Internet of which there are millions. To assist users in the exploration of the Internet are a number of publicly accessed facilities which allow organisations to both manage their own information and retrieve others. The most popular include Email, World Wide Web, Gopher, List Servers, Anonymous File Transfer Protocol (FTP), and Wide Area Information Server (WAIS).

WCMC is utilising these common facilities to provide information and metadatabase services as follows:

- network accessible disk at WCMC *(Anonymous FTP)*

- electronic forum for the CITES secretariat *(List Server)*

- *system support* for the *Microbial Strain Data Network.* This reflects WCMC's capacity to host external databases and Internet facilities.

The Centre is currently a **Beta-test** site for CIESIN interconnection software, and in conjunction with this is also installing a metadatabase facility for conservationists, ecologists and biologists. This will allow network users to browse through descriptions and view examples of WCMCs datasets. It is based on the *CIESIN catalog system* and lists WCMC data, including an inventory of its products and services. This belongs to the *NASA Global Change Master Directory* family of connected metadatabases.

In addition, the Centre has recently opened an experimental *World Wide Web* (WWW) server providing hypertext, text indexing and database queries of WCMC data. This provides WWW users with an attractive graphic interface within which they can interactively view text, maps and graphics on biodiversity and provides the Centre with global access to other centres of excellence, such as the Environmental Resource Information Network (ERIN) in Australia.

Initial experiences with this new advanced telecommunications technology have demonstrated great potential for the cooperative exchange of information, and for sharing of views, experiences, strategies and technology in a manner which is easily accessible to all.

4.4 SUMMARY

WCMC is an authoritative source of data on global biodiversity and conservation issues. Its aim is to support conservation and sustainable development programmes by providing comprehensive and up-to-date information and technical services, based on a programme of research and analysis. Through the cooperative use of networking, the Centre can disseminate information and ideas to enhance the quality of research in the field of environmental analysis and global change, and facilitate the exchange of scientific knowledge and technology required by the Convention on Biological Diversity.

5. CHALLENGES in the DESIGN and OPERATION of RETRIEVAL and REFERRAL SYSTEMS

5.1 TECHNOLOGICAL CHALLENGES

Section 2.2 identified the scope and some of the obvious difficulties of managing and exchanging biodiversity information in ways which are useful for strategy planning and priority setting. The *Information Revolution* has been underway for some years and information management tools have been developed to cope with the exponentially increasing quantity of data. Where appropriate, advanced technology can be applied to provide an effective clearing house function. The following sections summarise the current state and potential application of some of these evolving tools. Aspects of Information Technology such as telecommunications, database technology, mathematics and statistics, artificial intelligence are discussed.

Database Management Technology

It is obvious that environmental information for biodiversity policy decision-making will be voluminous, even given considerable summarisation and abstraction. Data management technology has had a number of recent advances of potential beneficial application. Until the mid 1970's database management systems (DBMS) were commonly either of a hierarchical or network structure. These data models are restrictive and so database structures lacked the flexibility needed by decision-makers.

Recent years have seen the established operational use of **relational database management systems** (RDBMS). RDBMS organise the information as a set of tables or **relations** linked by relational keys. Relationships between files are implicitly determined by common values in the **related** tables. Thus the RDBMS offers a flexible database structure which is easy to define, expand and change. Dynamic relations are created at query, permitting considerable exploration of relationships, typical of the needs of environment informational analysis.

Associated with RDBMS, has been the creation of **Structured Query Language** (SQL). This is an English-like retrieval language which adopts a mathematical set theory concept and uses the same symbols in query definition. This allows for easy use of various models independent of the brand-name of the RDBMS, thus considerable improvement in the potential for information integration. The standardisation introduced by SQL frees organisations from being committed to any particular brand of RDBMS and introduces the flexibility of SQL queries being issued across a distributed network of different RDBMS platforms.

The Object-Oriented Database Management System (OO-DBMS) is a relatively new approach which could give improved capability in dealing with groups of data items (objects) and connecting between objects. The OO approach provides an intuitive and powerful way of creating and implementing data models of complex systems such as ecosystems. This technology is immature and therefore suffers from poor performance and lack of standards as did RDBMS a decade ago. The policy decision maker can look to the future for the potential of this technology to allow for information compatibility on a broad base.

There have been recent developments, as well, in distributed database systems. A distributed database management system (DDBMS) is a single logical database that is physically distributed on several computers, communicating through a network. It provides multiple access to the database and mechanisms to avoid conflict in update, retrieval and backup of data. DDBMS can be very useful for very large databases where most users are concerned with only a small subset of the database, while permitting access to the user with broader needs. In reality, existing databases are hard to link in aDDBMS because of differing technology at each site. A solution to this which is now developing is the **Federated** database management system. It shares the advantages of the DDBMS, while additionally allowing data communication among DBMSs of different data models, brand-names and access methods. Each local DBMS is autonomous in that it preserves its own characteristics as viewed by the local user. The environmental policy decision support system of the future will no doubt use this technology, to link heterogeneous information collections in various countries and agencies. The technology is expected to mature in the next 5 to 10 years.

Geographic Information Systems Technology

Geographic Information Systems (GIS) technology has been used to manage environmental information and aid decision-makers since the inception of the Canada Geographic Information System (CGIS) in the mid 60s. The basic advantage of a GIS is its ability to manage and perform complex processing on the spatial component of the data as well as the statistical aspects. In that way the actual geographic boundaries of regions defined on the earth's surface can be manipulated. This allows for the integration and summarisation of environmental information using natural units -- such as watersheds, natural forest areas, soil units and so on, and to combine these effectively with man-made administrative data collection units. Thus it can provide the link between the decision-maker's viewpoint and the natural boundaries of the problem.

Several global data compilation projects, underway or recently completed, which are producing GIS compatible information are the Soil and Terrain Database (SOTER), the Digital Chart of the World (DCW), and Canada's Geoscope, to name a few. All of these are producing complete global coverages of generalised baseline digital data, ready for analysis and integration with thematic time-series.

In spite of the growing interest and demand, current GIS technology is limited in its applicability for environmental decision-making. First, regional spatial databases are frequently very large, and exceed the capacity of existing systems, and second, it is commonly necessary to store and process in excess of 100,000 spatial objects (eg; soil units) requiring millions of coordinate pairs or raster cells to define. These two factors -- sheer data volume, and the need to perform repeatedly complex geometrical computation on spatial objects - are the cause of the practical upper limit on the useable size of spatial databases in current GIS. Even with only moderately large databases, users commonly note "*lack of user friendliness*", "*slowness*" and "*unreliability*". Most GIS have little means of representing or processing information which is uncertain or of variable quality. A further restriction on usefulness is the lack of standardisation of internal structures and access methods. No standard spatial access protocol like SQL is yet on the horizon, so it may be some time before the heterogeneous distributed GIS needed for decision support is a reality.

Harmonisation Tools

Effectively combining information from different environmental sectors requires knowledge of the measurement techniques, standards, classification systems, quality assurance methods, and terminology used in obtaining and describing the information. This associated data or **metadata** must be maintained and organised for use in integration and harmonisation models on an ecosystem framework. Modern meta-databases are now using hypertext linkages and multi-lingual thesauri to form flexible relationships between the information descriptors. This allows for effective intelligent guidance to the decision maker on locating the appropriate information, its quality and accuracy, and its appropriateness to models. Such meta-databases are the key to an effective intelligent human interface to an environmental decision support system. UNEP-HEM's HEMDisk, the UNEP-GRID, WCMC and EEA metadatabases, and the NASA Master Directory are important beginnings in this regard.

Particularly in the area of ecology, much of the data has a subjective element, and considerable uncertainty. This results from the intrinsic complexity and randomness of nature, combined with finite budgets, and so can never be entirely removed. Biodiversity conservation decisions must always be made in the presence of imperfect data. The technology of uncertainty management is therefore an essential harmonisation tool. There have been recent advances in this area, employing fuzzy set theory and fuzzy logic to manage retrieval in databases with subjective or **verbal** attributes, and the use of spatial uncertainty zones and error propagation techniques to present visualisation of relative uncertainty to decision makers in the form of an uncertainty map.

Management and Integration of Text

The recording of biological and environmental phenomena still relies heavily on a descriptive (narrative) approach which can only be captured in free text form. For example, the description of a protected area is effectively achieved through narrative describing its floral and faunal characteristics.

The basic tool for this will continue to be the word processor. However, scientists have been reluctant in the past to rely too heavily on word processing documents as they are not as readily queried or updated as database tables. Therefore, their main use has been limited to reference material that is likely not to change (eg; site descriptions).

There have been software tools available for many years that have increased the utility of digital text through the application of text indexing. The commercial software product **Folio Views** is perhaps the most known and used, providing users with instant access to text, based on a word search. These types of tools are as flexible as the more popular database management systems, though the contextual form of free text can often be more informative. Such tools have been used very successfully in the past to retrieve metadata, reports and data.

With the advent of network communications there is going to be a renaissance in the use of narrative text as a form of information storage and retrieval. Network tools and services such as email, gopher, WAIS and WWW are ideally suited to the querying and retrieval of text documents. Tools such as WAIS are particularly powerful text indexing systems allowing Internet users to query distributed sites based on word searching.

The trend toward the use of WWW will make the integration of text and data more powerful as it encourages the use of text not just as a format to hold information and data, but also as a means of guiding users in the retrieval and use of data. For example, users of a WWW page may be selective in retrieving information based on related text describing sources and quality. These new tools coupled with the very large user base provided by the Internet have empowered organisations with large text based archives to become active in the dissemination of information and data.

Network Communications

As electronic networks become all pervasive and information sources become ever more comprehensive and sophisticated, one approach to reducing the complexity is through the establishment of Special Interest Networks (SINs). These `virtual' networks, focused on some particular theme, are embedded in the broader network infrastructure, but are used to communicate information relevant to that particular theme. The SINs function as a modern equivalent to learned societies, in that they perform a number of functions such as publication (journals, newsletters, datasets, software), library services (links to both on-site and off-site sources), communication (mailing lists, workshops and conferences), and services (specialist advice, data analysis). Structurally, a SIN consists of a series of participating nodes, one of which may act as a coordinating node. To become a node, a site must contribute to the SIN's activities and adopt its standards and protocols.

International databases such as Genbank are public compilations of contributions from thousands of scientists. A good example of a SIN is the European Molecular Biology Network (EMBNet), which serves the molecular biology and bio-technology research communities. It comprises nodes in numerous European countries.

Special Interest Networks are currently being implemented on the World Wide Web for biodiversity (for those familiar with the hieroglyphics of this it can be accessed through http://www.ftpt.br/bin21/bin21.html).

A variant on this theme is the establishment of **virtual libraries**. These are libraries managed by individual agencies but made accessible through electronic networks by collaborative agreement. The World Wide Web Virtual Library Subject Catalogue, maintained at CERN in Switzerland has topics ranging from Aeronautics and Aeronautical Engineering to Sport and including Biosciences. The Biosciences component of this virtual library is maintained at Harvard University in the United States.

Libraries, in the more widely accepted sense, have pioneered much of the thinking on networking information and have long developed collaborative arrangements for cataloguing, bibliographic searching and inter-library loans. Libraries and other agencies and programs are endeavouring to aid in the provision of information by providing bibliographies and directories to data sources, and tools to allow users to browse these directories (or metadatabases). As previously mentioned, UNEP-GRID and WCMC are examples of international agencies coordinating biodiversity information. National programs, some with major international dimensions include the US National Aeronautics and Space Administration (NASA), Global Change Master Directory (GCMD), the Consortium for International Earth Science Information Network (CIESIN) and, in the UK, GENIE.

The GCMD, is a good example of how a multi-node clearing house might operate. It offers a comprehensive source of information about worldwide Earth science data holdings, particularly satellite remote sensing data. The primary access to the GCMD is through Internet. The main centre is at the NASA/Goddard Space Flight Center in Maryland, USA, with other nodes at Frascati, Italy at the ESA/ESRIN Earthnet Program Office, at the National Space Development Agency Earth Observation Center in Japan, and the Canada Centre for Remote Sensing in Ottawa, Canada. Each node contains an exact copy of the GCMD database, which is updated automatically every two weeks.

Many agencies set up coordinating centres that link to more specialised information sources accessible via other agency information services. Examples on the World Wide Web include plant fossil records, and plant viruses. At the genetic level, Genbank is the database of all published nucleic acid sequences. It is accessible through the Internet and entries in the database can be searched for and retrieved by Accession Number, Description, Locus Name, Keywords, Source, Organism, Authors, and Title of the Journal Article.

As can be seen from just these few examples from the many national and international systems under development, it is unlikely, in the short term at least, that information about datasets will be coordinated in one or a restricted number of directories, but rather need a *network of networks*.

Other Evolving Tools

A number of other areas of information technology which are evolving rapidly may have potential benefit to environmental decision-making. The two most relevant are advances in telecommunications and in knowledge-based systems for modelling. The former now permits the transfer of very high volumes of digital data between locations globally. This permits in theory, a global distributed database, although none as yet has been implemented. In practice, at the least, it allows for theefficient movement of bulk information through a clearing house mechanism for integration and summarisation at the decision-making location.

Knowledge-based or **expert** systems are intended to imbue the decision-maker with the effective reasoning power of a large group of experts who individually have contributed their knowledge. The applications for the environmental decision-maker are both to help locate the appropriate information (where the complexity of the data is too great for any one person to "*know*") and to provide the expert reasoning in the decision-support models.

5.2 INSTITUTIONAL AND MANAGERIAL CHALLENGES

Quality and Ownership

One of the major problems facing a clearing house is the assurance of the quality of the information exchanged. This is an important function, but on the other hand maintenance of the desired independence, implies a need for a *hands-off* approach. Those best positioned to ensure the quality of the information are the original information *owners*. Custodianship is the means by which responsibility for a dataset is assigned to and accepted by the most appropriate agency. It provides a mechanism to ensure that each information holding is established, maintained and made available by that agency best able to do so. Every dataset should have one and only one custodian. Custodianship has responsibilities towards data acquisition, management and documentation, as well as rights to determine the conditions under which information is accessed and used.

All datasets required in environmental information systems are complex, thus their establishment and maintenance requires specialist knowledge. This specialist knowledge is required not only to acquire and maintain data, but also to document the data and advise clients as to its fitness for various potential uses.

The concept of custodianship also includes responsibility for modelling and analytical tools, including expert systems and decision support systems, information on bio-technology, regulation, standards, traditional uses and the like.

The most appropriate custodian for a dataset is likely to be that agency which meets one or more of the following criteria:

- has sole statutory responsibility for capture and maintenance of the information

- normally is the first to record changes to the information

- 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

The responsibilities of a custodian include the following:

- to define and maintain quality standards

- to organise the building of the database

- to update the dataset

- to ensure the continued integrity of the data

- to ensure appropriate access to the dataset

- to maintain documentation of the dataset

- to advise on appropriate uses of the dataset

The principle of data custodianship in information networks is that all data available to the network is *corporately owned*, in that many network users have a stake in those data. Designated custodians, however, have responsibilities to collect and maintain the data and also rights to regulate its access and use. Conditions may be attached to data on a network, for example data may be used for government decision making, public information or research purposes but not for any commercial purposes, at least without a specific licence to do so.

Custodians are responsible for these licensing arrangements, which can unfortunately become very complex. For example, some agencies in a network may have licences to use data that would be infringed if that data were passed through the network, but not if it were passed directly to certain other networked agencies. In other cases, certain agencies may be authorised to access data in raw form, whilst others may only be able to access more generalised versions. The issue becomes considerably more complex for a more proactive Clearing House structure, for example one with a stronger *brokerage* role in which the clearing house is actively seeking to link up partners who have not agreed a priori to a standard set of conditions.

Establishing Cooperative Networks

It is clear that with the custodial concept and the acknowledged need for the clearing house mechanism to be a linked network, that cooperative arrangements must be established. Success with this must rely on the principle of mutual benefit, which must be well stated and obvious to all participants. It must be clear that independent goals of the partners (whether countries or international agencies) must not be jeopardised by participation in the cooperative. Some of the potential difficulties stem from the reality that potential partners may be competing for the same sources of funding or have conflicting goals and ideals. A feeling of ownership over information may be particularly strong where it has been obtained only at great cost and with enormous quality control effort. This connects to the need for complex agreements or *licences* which will have to be administered by the central coordinating clearing house administration.

Legal Issues

The transfer of information - for instance on resources, on government strategies, policies and legislation - may impinge on legal and conceptual views of sovereignty and security. The CBD is careful to try to minimise this issue with such words as "*Recognising the sovereign rights of nations over their natural resources...*", but the effective exchange of information on biodiversity resources and related technology can only occur in an atmosphere of mutual trust. Again the principle of mutual benefit must over-ride concerns of misuse of the information for strategic or political purposes.

The transfer of bio-technology information and other enabling technology, including software may be restricted by copyright, patents and the like, and the *ownership* of technology may not always lie with countries, but with private sector companies. While the spirit of the CBD advocates fully open technology transfer, in this area, the open exchange of information will likely have to be restricted to a referral system with meta-data which enables interested parties to be connected for private negotiation.

While the details of the laws differ between countries, it is often the case that the provider of information - which turns out to be incorrect and causes harm as a result - may be held liable for the damage caused. This liability could fall on the Clearing House mechanism, especially in a case of negligence, where, for instance, no reasonable attempt was made to ensure quality, or poor information management practices allowed information to be corrupted. The establishment of a network of expert custodians coupled with quality management procedures (eg; meeting ISO-9000) in the Clearing House mechanism would likely eliminate any possibility of negligence, and greatly reduce the chances of being considered liable. The **independence** of the clearing house is also an important factor in alleviating these concerns.

Operational Management

An information retrieval and referral Clearing House can operate with a relatively small administrative staff, particularly if it is established as a network of custodians with a central coordinating function. The coordinating node will be engaged in the following as a minimum:

- negotiating custodial agreements

- drafting terms and conditions or a protocol for cooperation

- soliciting cooperation from a network of clearing houses

- managing the technology of information retrieval and exchange

- managing and controlling thesauri and meta-databases

- providing advice and instruction on how to access information through the network

These functions would best be carried out by an established organisation which has an international information exchange mandate and experience with the technology of networking and information management.

It is important to note that the operational management requirements of brokerage, technology and policy services are much more complex - involving the need for scientific and policy support staff and technological experts covering a very wide range of technology.

6. THE WAY FORWARD: A COOPERATIVE CLEARING HOUSE NETWORK

6.1THE PRINCIPLES

The principles for a successful cooperative Clearing House were well stated in the recommendations of the second session of the Intergovernmental Committee and in the background paper. They can be paraphrased as follows:

- an openly accessible electronic data network

- a decentralised network of national and regional centres with central coordination (for example in a United Nations agency)

- based as far as possible on existing institutions

- use as much as possible existing databases, information services, and networking services

- driven by the real needs of the Contracting Parties

- small in the initial phase but with the capacity to be further developed.

6.2 THE CONCEPT

To achieve these principles, the Clearing House conceptually should be a linked electronic network with access through a central focal point or **hub**. The nodes in the network would be, where possible, existing more specialised clearing houses in the various sectoral areas required to cover the range of information and technology exchange specified or implied by the Convention.

Two *principles* of 6.1 which impinge directly on the conceptual model are:

open accessibility which implies:

- the use of widely accessible networking systems such as Internet

- use of public domain information searching and access tools

- central or collective payment of costs, so that cost is not a factor limiting access for developing countries

small in the initial phase which implies:

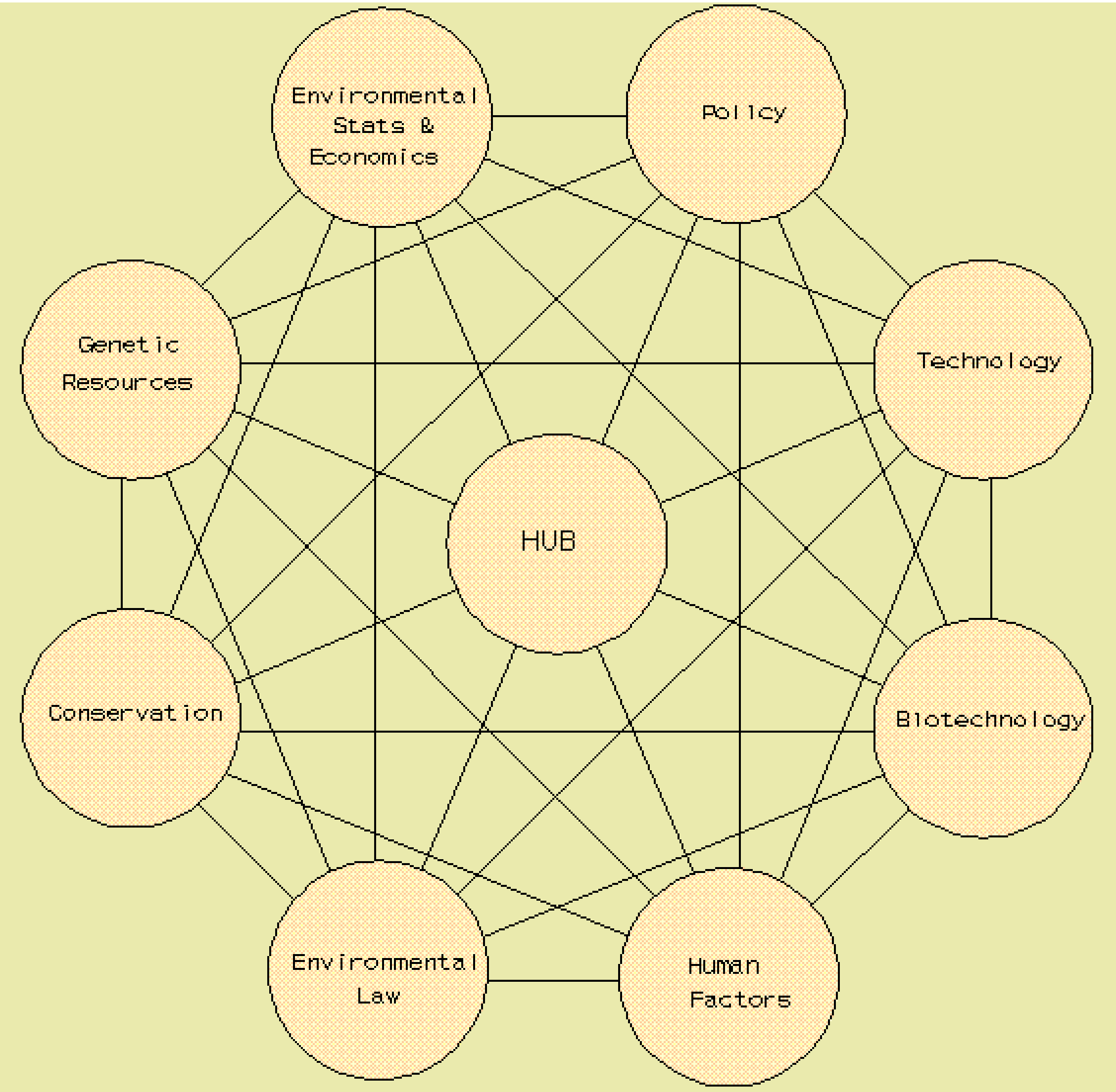
- the first phase of the Clearing House might mainly be a referral system, that is, concentrate on meta-databases and associated thesaurus and searching tools

- the policy and technology aspects might initially be accommodated as information clearing house nodes, not as separate pro-active services or functions

- the brokerage service might be considered at a later date, when the information exchange functions of the clearing house are well established.

One approach would be to have principal nodes (in the first phase) for each of the eight major suggested thematic areas listed in Section 1.1.

Figure 1:*A conceptual view of a cooperative clearing house network incorporating these thematic nodes.*



6.3 PRACTICAL CONSIDERATIONS

Given the above principles and concept, there are some practical considerations regarding the establishment of a clearing house mechanism.

First, as has been noted elsewhere in this document, it is important for the clearing house to concentrate initially on facilitating information exchange as this is a difficult enough task in itself. It will be important to establish early all of the following:

- identified real needs of the Convention

- clear mission for the Hub organisation

- resource estimates for the HUB and network in general

- proforma agreements with terms and conditions for cooperating nodes

- meta-database structures

- access methods - as open and *low-tech* as possible

- complimentary vocabulary control (thesauri)

- quality management principles (for the nodes)

- quality management practices (for the Hub)

- cost sharing mechanism

- internal operational procedures for the Hub

Secondly, to achieve these initial goals for a Clearing House, as much advantage should be taken as possible of existing information exchange standards, metadatabases and thesauri. In this regard, particular attention should be paid to the following:

- metadatabase and format of the NASA Master Global Change Directory

- UNEP-GRID metadatabase

- EA Catalogue of Data Sources

- WCMC metadatabase definition

- CODATA thesauri

- INFOTERRA Thesaurus of Environmental Terms

- ERIN (Australia) metadatabase

- OECD environmental reporting formats

- UNEP Guidelines for Country Studies

- Biodiversity Metadatabase for East Africa

- Metadatabases and guidelines of GCOS and GOOS

- Guidelines for Biodiversity Information Management (currently under development by WCMC for UNEP-GEF)

Finally, to facilitate the rapid start of a Clearing House, consideration should be given to identifying some agencies who would volunteer to be initial nodes (who are currently active clearing houses) and appoint one as **Hub** on an interim basis while facilities are being developed for a permanent Hub structure.

LIST of ACRONYMS CBD Convention on Biological Diversity CIESIN Consortium for International Earth Science Information Network
CITES Convention on International Trade in Endangered Species
DBMS DataBase Management System
DCW Digital Chart of the World
ECDIN Environmental Chemicals Data and Information

Network (of the EC)
EEA European Environment Agency
ELC Environmental Law Centre
ELIS Environmental Law Information System
EMBNet European Molecular Biology Network
EPA U.S. Environmental Protection Agency
ERIN Environmental Resource Information Network (Australia)
FAO Food and Agriculture Organisation
FTP File Transfer Protocol
GCMD Global Change Master Directory
GENIE Global Environmental Network Information Exchange
GIS Geographic Information System
GLIS Global Land Information System
HDGECP Human Dimensions of Global Environmental Change Programme
HEM Harmonisation of Environmental Measurement
ICSU International Council of Scintific Unions
INET 2000 National Telecommunications Network
IGBP International Geosphere-Biosphere Programme
IOC Intergovernmental Oceanographic Commission
IPGRI International Plant Genetics Resources Institute
IRPTC International Register on Potentially Toxic Chemicals
IUCN International Union for Conservation of Nature and Natural Resouces
MRRL Midlands Regional Research Laboratory
NASA National Aeronautics and Space Administration (U.S.)
SDN Sustainable Development Network (of UNDP)
SIN Special Interest Network
SOTER Soil and Terrain Database
SQL Structured Query Language
UNCED United Nations Conference on Environment and Development
UNDP United Nations Development Program
UNEP United Nations Environment Programme
UNESCO United Nations Educational, Scientific and Cultural

Organisation
UNIDO United Nations Industrial Development Organisation
USDA U.S. Department of Agriculture
USGS U.S. Geological Survey
WAIS Wide Area Information Server
WCMC World Conservation Monitoring Centre
WCRP World Climate Research Programme
WDC World Data Centers
WDCGG WDC for Greenhouse Gases
WDCM World Data Center for Collections of Microorganisms
WFCC World Federation for Culture Collections
WHO World Health Organisation
WWF World Wide Fund for Nature
WWW World Wide Web

ACKNOWLEDGEMENTS

This Background Paper was compiled by Ian Crain, Senior Advisor Biodiversity Information Management, World Conservation Monitoring Centre, with input from Ian Barnes, Mark Collins, Gareth Lloyd, Gwynneth Martin, Jeremy Harrison, Donald Gordon, Jo Taylor and Laura Battlebury of the Centre, and John Busby and Arthur Chapman of ERIN. The outlines of the roles and activities of other clearing house agencies have, in part, been abstracted and edited from public domain sources such as UNEP's HEMDisk, World Wide Web, and published documents. Background research has been supported by the UK Department of Trade and Industry.
