On networking Indian biodiversity databases

"A little information, when shared, can go a long way!"

A.K. Chakravarthy, Anand S K, Arundhati Das, D.K. Bhaskar, Gangadhar V. Maddikery, H. Nagaraj, Harish Bhat, Harish Chandra Karnatak, Indrani Chandrashekar, J. R. B. Alfred, Jayant R. Haritsa, K. Gopinath, K.V. Gururaja, Karthik Shankar, L. Shyamal, M.B. Krishna, Madhav Gadgil, Meenakshi Munshi, Mohammed Irfanullah, N. R. Menon, Narasimha Murthy, P. N. Krishnan, P.R. Seshagiri Rao, P.S. Roy, Pramod Subbarao, R. Keshavachandran, R. Sundar Raj, Raghavendra Gadagkar, Rekha, Renu Swarup, S. Karthikeyan, S. Krishnan, Sandeep Sharma, Santosh J. Eapen, Sarnam Singh, Sathyanarayana Bhat, Shivanna, IFS, Srinidhi. S, Subbiah Arunachalam, Suja, Sundeep Sarin, T.B. Rajashekar, T.V. Ramachandra, V. V. Sivan, V.K. Gupta, V.V. Ramamurthy, Vasu, G.J., Vijay Barve, Vijay Edlabadkar, Vishwas Chavan

1. The Rationale

India is one of the world's top twelve megadiversity countries, rich in biodiversity resources. We also possess a wealth of knowledge associated with biodiversity, be it the orally held knowledge of folk healers or herders, or the traditional knowledge codified in Ayurvedic, Sidha or Yunani texts. Its biodiversity resources are far better known scientifically than those of other tropical megadiversity countries such as Brazil or Indonesia. As a result, India has developed a number of excellent biodiversity databases such as the Flora of Karnataka (Ganeshaiah, et. al., 2002), Traditional Knowledge Digital Library (NISCAIR, 2002) or the National Register of Green Grassroots Innovations and Traditional Knowledge (NIF, 2002). There exist therefore rich possibilities of building upon country's biodiversity resources and associated knowledge; to promote biodiversity-based enterprises in the modern, as well as traditional sectors; to develop biotechnology industries at the cutting edge of new technologies as well as to encourage local level value addition to biodiversity resources. Important new markets are also emerging for produce of organic agriculture. Taking advantage of these markets will require development of good databases on agro-ecosystems, including incidence of pests and diseases.

India has recently passed a pioneering piece of legislation in the Biological Diversity Act 2002 that provides a framework for taking advantage of several significant new provisions of the international Convention on Biological Diversity (CBD):

1. Assertion of sovereign rights of India as a country of origin.
2. Assertion of intellectual property rights over codified traditional knowledge like Ayurveda and Yunani systems of medicine.
3. Assertion of intellectual property rights over orally transmitted traditional knowledge.
4. Assertion of intellectual property rights of grass-roots innovators.

The Biological Diversity Act visualizes the establishment of a National Biodiversity Authority (NBA), State Biodiversity Boards (SBB) and Biodiversity Management Committees (BMC) at the level of all local bodies, namely, Gram, Taluk and Zilla Panchayats, as well as Municipalities and Corporations. The NBA working with SBBs and BMCs will have the responsibility for and authority to:

(1) Decide upon the admissibility of all patent and other Intellectual Property Rights (IPR) applications based on Indian biodiversity resources and associated knowledge in consultation with relevant local Biodiversity Management Committees.

(2) Decide upon applications to access biodiversity resources and associated knowledge for commercial use in consultation with relevant local Biodiversity Management Committees.
(3) Decide upon appropriate benefit sharing arrangements in relation to IPR applications in consultation with relevant local Biodiversity Management Committees.
(4) Issue guidelines on appropriate collection fees and other benefit sharing arrangements in relation to applications to access biodiversity resources and associated knowledge for commercial use in consultation with relevant local Biodiversity Management Committees.
(5) Decide on admissibility of joint research proposals involving foreign agencies.
(6) Decide on priorities and appropriate actions for conservation and sustainable use of natural populations of biodiversity resources.
(7) Decide on priorities and appropriate actions for maintenance of health of natural ecosystems.
(8) Decide on priorities and appropriate actions for conservation of domesticated biodiversity.
(9) Decide on priorities and appropriate actions for constitution of heritage sites.
(10) Decide on priorities and appropriate actions for identification of threatened species.
(11) Promote scientific research pertaining to biodiversity and associated knowledge.
(12) Promote public awareness pertaining to biodiversity and associated knowledge.

Evidently, the NBA, SBBs and BMCs would need a well-organized information system on India's biodiversity resources and associated modern as well as traditional, codified as well as oral knowledge to do justice to these ambitious objectives. Such a system will have to deal with a whole range of spatial scales from local to national as well as link properly with global databases. It will also have to address issues such as linking to information on the large holdings of biological specimens of Indian origin located in herbaria and museums abroad.

India with its emerging strengths in Information Technology (IT) as well as biotechnology is in an excellent position to turn this array of significant challenges into welcome opportunities. This calls for networking of country's existing biodiversity databases to take advantage of synergies, and to link all of these to activities leading to value addition. As a part of this process, the existing biodiversity databases will need to be considerably augmented and strengthened, and new ones created. We will have to come up with novel ways of bringing on board the substantial knowledge base of country's barefoot ecologists and grass-roots innovators. We also need to devise a country wide decentralized system of monitoring biodiversity. Such a decentralized system could serve to enhance the quality of education by engaging teachers and students in first hand understanding of biodiversity and associated knowledge and in creating, using, and managing electronic databases, including those employing Indian languages.

2. Current Scenario

A wealth of information exists on India's biodiversity resources and associated knowledge. This may be in form of specimens, gray literature such as unpublished reports of

District Floras project or Forest Working Plans, and books, monographs and scientific papers. A good beginning has been made in organizing a part of this information in the form of electronic databases. Some of the key initiatives in this context include:

- Agricultural Databases and information on sacred groves (MSSRF, 2003),
- Agricultural Research Information Network (ARISNET)(Sreenivasulu and Nandwana, 2001)
- Bibliographic and referral information on Western Ghats (CES, 2003).
- Biodiversity characterization using RS/GIS (Roy and Ravan, 1996; Roy and Tomar, 2000; and Roy, et.al., 2002),
- Biotechnology Information System (BTISNet) (DBT, 2003),
- CDROMs on Marine Prawns, Marine Crabs, Mangroves, Lignicolous Fungi and corals of India (NIO, 2003),
- Endemic Trees of Western Ghats (Datta, et.al. 1997),
- Environmental Information System (ENVIS) (MoEF, 2003),
- Ethnobotany (NBRI)
- Flora of Karnataka (Ganeshaiah, et.al., 2002),
- LIFKEY/LIFDAT (CES,2003)
- Indian Medicinal Plants database (FRLHT, 2003),
- National and State Forest Vegetation maps and National Basic Forest Inventory (NBFIS) (FSI, 2003),
- National Register of Green Grassroots Innovations and Traditional Knowledge (NIF, 2002)
- National Wildlife Database and Zoo Database (WII, 2003),
- NCL Center for Biodiversity Informatics (NCL, 2003), Birds of India (SACON, 2003),
- People's Biodiversity Registers (CES, 2003).
- Plants of India and Legume Database of South Asia (NBRI, 2003),
- SAHYADRI: Western Ghats Biodiversity Information System (database of Western Ghats flora and Fauna (http://ces.iisc.ernet.in/biodiversity)
- Sasya Sahyadri (Ganeshaiah, 2003),
- Traditional Knowledge Digital Library (NISCAIR, 2002)

There are, however, many lacunae and we may summarize the situation as follows:

1. Many institutions have information documented in the form of databases.
2. Databases are in heterogeneous formats.
3. Few are on the web, while many are available offline.
4. Some of these are well-structured, others are largely project /species specific and/or unstructured.
5. These databases exist independently.
6. There is no framework to link the scattered data so as to facilitate exchange of data amongst the different databases.
7. There is no meta-data.
8. The gap between data managers and data producers is widening.

3. Institutional and legislative framework

The objectives that need to be addressed range over facilitating development of new drugs based on knowledge of folk healers or that in Ayurvedic texts, and conservation of endemic species and of land races of domesticated animals, to promoting sustainable harvests of non-timber forest produce, and equitable sharing of benefits with grass-roots innovators. Manifestly, a variety of data, spatial and non-spatial, on a diversity of scales (global, national, state, district, village or biogeographic provinces, biomes, landscapes and landscape elements), belonging to diverse knowledge systems (modern science, Ayurveda or Yunani, knowledge of folk healers, farmers, fisherfolk, herders, tribals) and in multiplicity of languages needs to be brought together and organized to meet these objectives.

We clearly need well thought out institutional arrangements and legal provisions at the national level, as well as in terms of links with international agencies to address these manifold concerns. The Department of Biotechnology, GoI, with its National Bioresources Development Board has thus far provided the lead in organizing relevant activities. The Ministry of Environment and Forests, GoI, which would soon establish the National Biodiversity Authority will clearly come to play a significant role in the coming days. A number of other Governmental agencies such as Ministries of AYUSH, Health and Commerce, CSIR, ICAR, Department of Space, National Innovation Foundation, State Biodiversity Boards, as well as the many Universities and NGOs would need to work together for this purpose. These should constitute a network coordinated by some nodal agency such as the National Bioresources Development Board with a clear delineation of the specific function of each institution. These need to interact, as appropriate, with international agencies like CBD, Global Environment Facility, WTO, WIPO-IGC, CIPGR, UPOV, UNCTAD and the World Bank.

Appropriate statutory agreements, including Information Transfer Agreements (ITA) and Material Transfer Agreements (MTA), supported by existing legislation or legislation that may have to be specially developed, will have to be put in place to specify ownership as well as benefit sharing arrangements while establishing linkages and organizing exchanges amongst the different constituents of the network. This is particularly relevant in the context of intellectual property rights including those of holders of traditional knowledge and grass-roots innovators. Finally, national level metadata specifying the content of the constituent databases would have to be built up.

5. The Challenge

The challenge before us is to set standards and make technological choices that would facilitate networking of databases, and add real value to the information being brought together, while at the same time, (a) maintain the autonomy of the various databases and ensure that there is abundant scope for expression of creativity and originality of the designers and managers of different databases, as also (b) ensure the security of the data and (c) protect all legitimate intellectual property rights. This would obviously have to be worked out as a group exercise by all concerned institutions and individuals. A first step would have to involve (a) an inventory of the on-going Indian efforts, and (b) a review of the various pertinent standards, technologies and protocols developed anywhere in the world. These surveys would have to address issues of data (i) characterization and classification, (ii)

validation and authentication, (iii) organization and structuring, (iv) storage, archival, warehousing, (v) retrieval, (vi) dissemination, (vii) sharing and interoperability, (viii) access, (ix) security, and (x) visualization, analysis and value addition, as well as (xi) use of multiple languages, and (xii) capacity building needs.

5.1 Entities

While we did not have adequate time at the Workshop to deal with the large number of issues that would have to be addressed, we would like to offer a few preliminary suggestions. Firstly, the scope of the databases that need to be thus brought together would have to go beyond the taxon-centric databases that is the exclusive focus of many efforts including the Global Biodiversity Information Facility (GBIF). Thus we visualize linking of data on medicinal uses and chemical composition to taxonomic data on medicinal plants. We also need to link taxonomic data on medicinal plants to data on geographical distribution, abundance, and harvest levels on land under different forms of ownership and access regulations to support development of strategies for conservation and sustainable use of these plant populations. We therefore suggest that the networking effort brings under its purview databases that will deal with the whole range of categories of entities listed in Table 1.

Table 1: A possible framework for definition of entities for Biodiversity Databases

Class of entities	Sub-class	Examples
Geographic	Physical	Mountain ranges, river basins, wetlands
	Political	Local bodies, states, UTs, GoI
	Ecological	Agro-ecological zones, Biomes, Landscape elements
	Management regimes	Reserved forests, wild life sanctuaries, sacred groves
Biological units	Genes	Bt genes in Bt cotton
	Taxonomic groups	Species, orders
	Natural / cultured	Varieties registered under Protection of Plant Varieties and Farmers' Rights Act
	Management status	Species listed under Schedules of Wild Life Act, CITES, or in Red Data Books, sacred species
People	Political units	Citizens of India, citizens of particular states or Panchayats,
	Membership of organizations	University faculties, Staff of R and D labs
	Occupational groups	Hakims, Vaids, Biotechnologists
	Social groups	Indian communities as recognized by Anthropological Survey, Scheduled tribes
	Holders of material property rights	Holders of rights of collection of forest produce from particular localities
	Holders of	Holders of patents

	intellectual property rights	
	Linguistic units	Speakers of different languages and dialects
Organizations	Relation to government	Government agencies, NGOs, Inter-governmental agencies
	Objective	Commercial, Not-for-profit, Scientific
	Nationality	Indian, foreign
Biological populations	Population levels	Population levels of particular species in specific localities
	Manipulations	Harvests or production under cultivation, including specific techniques employed, of particular produce of particular species in specific localities
	Transport	Transport of produce of particular species in specific localities to other specific localities
	Marketing	Marketing of produce of particular species in specific localities in other specific localities
Biological materials	Processing	Preparation of acetone extracts from particular plant species
	Value addition	Preparation of plant based drugs or cosmetics
	Technology	Stabilization of alkaloids extracted from Neem
	End products	Particular molecules isolated from biological sources
	Uses/disservices	Therapeutic, cosmetic, allergenic
Knowledge	Nature of knowledge	Satellite imageries, folk taxonomies, medicinal properties of particular species
	Source of knowledge	Scientific research, Classical tradition, Individual hakim
	Medium	Scientific journals, palm-leaf manuscripts, oral traditions
	Generation of knowledge	New collaborative research, new Indian research, grass-roots innovations
	Transmission of knowledge	Web-sites, scientific publications, scientific meetings, orally from mother to daughter
	Access to knowledge	Public domain, patented, trade secrets
	Rights over knowledge	Individual, community, corporate
	Validation	Raw, Validated through different techniques
Disputes	Disputes over access to material or knowledge resources	Refusal of permission to an application for collaborative research

5.2 Architecture

Of course, there are many useful lessons from the on-going international exercises for us. Thus, the International System Of Patent Classification might serve as a useful starting point for a system of classification of database entities. One may particularly mention here the Traditional Knowledge Resource Classification being developed as a component of this system as a result of Indian inputs. We might also with profit build upon the Data Architectural Model of the Global Biodiversity Information Facility (GBIF). GBIF intends to make world's biodiversity information available to all within the next 10 year period. Employing the architectural model summarized in Table 2, the GBIF is currently serving over 10 million records from 34 distributed databases.

Table 2: GBIF Web Services Architecture

Registry services	UDDI
Interface description	WSDL
Access protocols	SOAP, DiGIR
Data encoding	XML, XML Schema
Transport	HTTP over TCP/IP

5.3 Data Quality

Data quality is clearly an overriding concern. A whole series of standards will have to be developed and checks organized at the level of data creation, organization and sharing to ensure high quality. Here one may mention a modern tool that may be employed in connection with data meant to be publicly available, namely, ***Wiki-wiki*** pages. These web-pages permit any visitor to edit portions designated as open to editing. This permits all interested parties to correct mistakes and add to the information. Of course it is possible that some visitors may maliciously distort the material. However, this is not a serious problem since the earlier versions can be preserved and reinstated in case of such a mischief. The experience of *Wikipedia,* a public knowledge resource encyclopedia has been very positive. Portions of the biodiversity database may therefore be thus maintained and further developed as *Wiki* pages.

5.4 Confidentiality

Another significant concern relates to the need to maintain confidentiality of some of the data. This may be appropriate in the following contexts:

1. Results of scientific research of potential commercial applications.
2. Codified traditional knowledge such as of Ayurveda that is publicly available, but may be converted into a value added product such as Traditional Knowledge Digital Library.
3. Oral traditional knowledge or grass-roots innovations, thus emanating from sources outside modern scientific tradition, of potential commercial applications.
4. Information on occurrences of biodiversity resources that are potentially exposed to threats of over-exploitation.

Information on issues (3) and (4) may be provided by knowledge holders who are outside the modern scientific, or even literate tradition, and therefore may require the involvement of an agency to interface with the world of science, technology and commerce. The current tradition in disciplines like anthropology and ethnobiology fails to

provide any credit to grass-roots knowledge holders. We need to change this system and ensure that such knowledge holders get (a) full credit for the knowledge they may make available with an understanding that it may be made public, and (b) full protection for the knowledge that is provided with an understanding that it may be kept confidential and made available only to certain parties under specified conditions. Clearly, a new type of agency is needed to play such a role. NIF is a possible candidate for this role. It is suggested that it may serve as a repository of and facilitator in value addition to such knowledge. It would interact with the knowledge holders, either individuals or communities through the medium of Memoranda of Agreements. Such MOAs would take place of the generally recommended PIC (Prior Informed Consent), since the PIC is a one-way transaction and does not incorporate an element of reciprocal responsibility on part of the agency receiving the information(see Figure 1).

Figure 1 presents a possible model of dealing with grass-roots knowledge based on discussion at the meeting of the Governing Council of NIF on March 9, 2004.

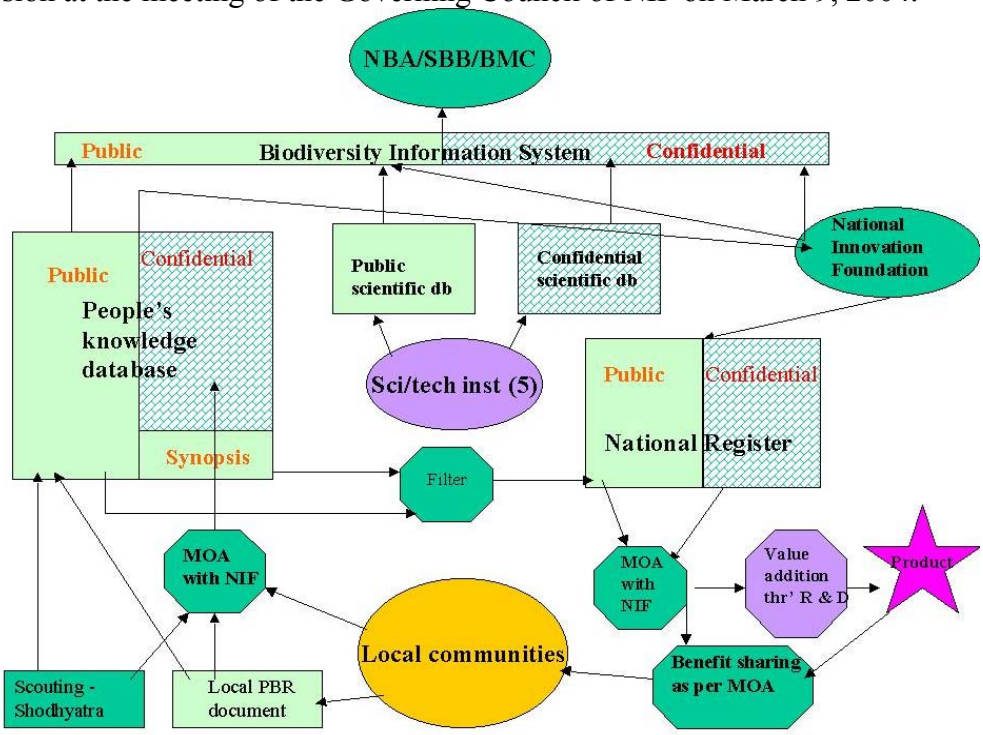

Figure 1: One possible model of dealing with public and confidential components of knowledge associated with biodiversity.

BMC: Biodiversity Management Committees at the level of all local bodies, namely, Gram, Taluk and Zilla Panchayats, as well as Municipalities and Corporations

Filter: People's Knowledge Database will include all information recorded by people, either as public or confidential in the PBRs. Part of this information considered to be novel and socially and environmentally acceptable will be incorporated in the NIF's National Register. Filter refers to the criteria used to decide on which information will be incorporated in the National Register.

MOA: Memorandum of Agreement between knowledge providers and the agency, such as National Innovation Foundation serving as a repository of and facilitator in value addition to the knowledge.

NBA: National Biodiversity Authority

NIF: National Innovation Foundation, an agency established by the Government of India to reward and promote traditional knowledge and grass-roots innovations.

PBR: People's Biodiversity Registers, documents incorporating information on biodiversity and associated knowledge at the level of local bodies (Panchayats and Municipalities) as specified in the proposed rules for Biological Diversity Act.

SBB: State Biodiversity Boards

Shodhayatra: A march through the countryside organized by National Innovation Foundation to document traditional knowledge and grass-roots innovations.

Synopsis: An index of confidential information made public to provide an indication of the content for possible value addition without disclosing the full information.

6. A Road Map

We believe that our ultimate aim should be to provide mobile, multi-lingual biodiversity information to anyone, anytime, anyplace. To move towards this goal we suggest the following steps:

- Prepare an inventory of Datasets
  - Who has what?
- Prepare a meta-database of all pertinent databases
- Assess the complementarity of the different databases and identify major gaps
- Review various biodiversity data standards, protocols and technology
- Promote and facilitate digitization of non-electronic data (biological specimens, legacy literature, etc.).
- Address unstructured data management and dissemination issues
- Set standards for information gathering for various entities, especially in contexts of ***People's Biodiversity Registers***
  - Minimum set of information that should be gathered
- Devise a list of entities for which such standards can be proposed
  - E.g. Species, Landscape elements
- Organize a system of public, transparent scrutiny and elimination of errors in the datasets
- Work out shared conventions for definition of entities and relationships amongst entities
- Adopt / evolve standards, protocols, and technology for linking Indian biodiversity databases
- Develop mechanisms /protocols for data sharing
  - Standards for data access, authentication / validation and security.
  - MOU? (Between institutions, between projects sponsored by DBT or other agencies, etc.)
  - Resolve issues of ownership of shared information
- Organize multi-lingual data acquisition and dissemination
  - Decide on common conventions for Indian language applications: ISCII? Or Unicode?
- Devise standards and tools for data archival, rescue, data warehousing and data mining
- Develop analytical, visualization, virtual ecosystems, interactive, decision-support tools development
- Organize, in parallel with other activities, activities geared to capacity building

❖ Evolve federated mechanism for access to these heterogeneous data sources preferably through single data portal

7. Short-term Milestones

We would like to suggest that the following activities be immediately initiated.

- Inventory / Survey of existing databases – 3 months
  - Questionnaire development: 3 weeks
  - Feedback on questionnaire: 3 weeks
  - Analysis of the feedback: 4 weeks
  - Action: Mr. Vijay Barve (FRLHT)

  A draft of this questionnaire is provided below in Box 1.
- Pilot for linking databases – 6 months
  - Identification of databases: 4 weeks

  (e.g. CES, ZSI, IIRS, NCL, etc.)
  - Resource planning: 4 weeks
  - Implementation: 8 weeks
  - Documentation of issues/ bottlenecks: 4 weeks
  - Action: Mr. Vishwas Chavan

❖ Pilot for multi-lingual dissemination – 6 months

Action: Cyberscape- Anand / CES

Outline of the Technical Data Collection Format

Technical Contact person from the organization with contact details

Details of the Database

- Name of the database
- Objectives and aims of the database
- Brief description

  Mechanism adopted in data compilation and digitization
  Details of data (primary / secondary/ ..)
  Data source

- Indicate the mechanism adopted in data validation -
- Stakeholders of the database
- Country, institutions, individuals providing the data
- Other sources of data
- Metadata details

Kinds of entities represented in the database:

Class of entities	Sub-classes

Geographical	Physical, Political, Ecological, Management regimes
Biological units	Genes, Taxonomic groups, Natural / cultured, Management status,
People	Political units, Membership of organizations, Occupational groups, Social groups, Holders of material property rights, Holders of intellectual property rights, Linguistic units
Organizations	Relation to government, Objective, Nationality
Biological populations	Population levels, Manipulations, Transport, Marketing
Biological materials	Processing, Value addition, Technology, End products, Uses/disservices
Knowledge	Nature of knowledge, Source of knowledge, Medium, Generation of knowledge, Transmission of knowledge, Access to knowledge, Rights over knowledge, Validation
Disputes	Disputes over access to material resources, Disputes over access to knowledge

- Priorities addressed by the database (taxonomic, political, conservation, policy making, socio-economic, others pl. add)
- Targeted user groups (policy makers, scientific community, educators, private company, general public, *students,* others pl. add)
- Contractual agreements in place or desired for data sharing/output/ acquisition (formal contract, MOU, letter of agreement, verbal contract, others pl. specify)
- How IPR, Copyright and financial gain issues are handled

- Source of funding
- Total cost incurred on database so far and projected total cost to complete the database
- Problems and hurdles faced / expected

Technical Specifications

- Platform and Technology used (Windows/Linux, MySQL/Oracle/XML...) with Version of each
- It is already web enabled in some format?

Provide URL details (and mode of access)

- Data types incorporated (text, dbf, mdb, xls, multimedia formats, others pl. specify)
- Data transfer methods (CD/Print Output/ Email/ Internet/ Interactive web search, others pl. add)

- Details of the database like Entities, Relationships, Structure, Fields, etc.

  Number of data tables

  Enclose E-R Diagram and Tables with relationships

- Method of data capture and data entry e.g.
  a. If data captured by field workers, if validated by others and then entered
  b. Accepted from literature
  c. Entered by hand-held gadgets / through user interface etc….

- Current access mechanisms
- Size of Database, Number of records, field densities

- If data has temporal dimension i.e. data may change with time. If so whether old data is preserved with time tag.
- Procedure followed in editing data, who is authorized , if pre-edited copy is retained etc

- Code sets used in storing database
- Presence of multilingual component and details
- Standards being followed
- Time scale for developing the database and completion of the database
- Estimated cost and time / record
- Special Remarks

Feasibility and willingness to participate in "Network of Indian Biodiversity Databases"

- Exchange / sharing data is feasible
- Exchange formats / standards adopted (xml, z39.50, CORBA, others pl. specify)

- Level of access restrictions i.e. if access restrictions are imposed then its level : table / record / field

- Details and contact information of other related databases you are collaborating with

List of key publications, reports and websites resulting from Database.

Acknowledgements:

This report is a product of a “Workshop on Networking Indian Biodiversity Databases” sponsored by the Department of Biotechnology, Government of India at the Centre for Ecological Sciences, Indian Institute of Science, Bangalore on March 23-24, 2004. The Centre for Ecological Sciences also wishes to acknowledge the financial support of the Ministry of Environment and Forests, Government of India.

References

To be added