CHAPTER 26

DATA NETWORKING FOR CONSERVATION BIOLOGY IN ASIA

By

LINDA OLSVIG-WHITTAKER¹, MEREDITH A. LANE²,
CHARLOTTE MACALISTER³, YAN XIE⁴ AND ANDREW T. SMITH⁵

¹ Science Division, Israel Nature and Parks Authority, 3 Am Ve’ Olamo, Jerusalem 95463, Israel.

² Global Biodiversity Information Facility, Universitetsparken 15, 2100 Copenhagen, Denmark.

³ Environment Programme, Mekong River Commission Secretariat, P.O. Box 6101, Unit 18 Ban Sihane Neua, Sikhottabong District, Vientiane 01000, Lao PDR.

⁴ Institute of Zoology, Chinese Academy of Sciences, Beijing, People’s Republic of China.

⁵ School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA.

ABSTRACT

We provide a snapshot overview of developments in data networks for conservation biology in Asia. Three main types of networks are described: special interest, regional and global. Although they overlap, each has characteristic goals and features. Special interest networks tend to be complex with many kinds of information, regional networks are intermediate in complexity, and global networks may be complex and limited in record capacity or relatively simple in design with very large databases. We give representative examples of each. Since this paper grew from a Society for Conservation Biology workshop held in Nepal, the focus is on networks with potential use for Asian conservation efforts. We report the rapid development of networks, their trends towards dispersed rather than centralized systems, and the necessity for professional consensus and trust in their development.

Keywords: bioinformatics, conservation, data management, internet, networking
INTRODUCTION

The ability to share information among individuals, groups and nations is an increasingly recognized need in conservation. For example, in 2005 the online journal *Biodiversity Informatics* was launched (Biodiversity Research Center 2006) and in 2006 Elsevier started publication of the online journal *Ecological Informatics* (Elsevier 2006). Conservation-oriented databases are multiplying rapidly: a quick Google web search with the terms +conservation, +database, +network resulted in 16,000,000 hits. This is clearly an idea whose time has come. The signing of the Convention on Biological Diversity (Secretariat of the Convention on Biological Diversity, 1994; Glowka et al. 1994) and its requirements (Articles 17 and 18) gives an additional thrust to development of various networks for sharing information and expertise. Article 17 of the Convention states:

**Article 17. Exchange of Information:**

1. *The Contracting Parties shall facilitate the exchange of information, from all publicly available sources, relevant to the conservation and sustainable use of biological diversity, taking into account the special needs of developing countries.*

2. *Such exchange of information shall include exchange of results of technical, scientific and socio-economic research, as well as information on training and surveying programmes, specialized knowledge, indigenous and traditional knowledge as such and in combination with the technologies referred to in Article 16, paragraph 1. It shall also, where feasible, include repatriation of information.*

During the intervening ten years, many international conservation information networks were developed and came online (See Table 1). Most are web-based and open to the public, serving a variety of users. Conservationists, whether managers or planners, usually want information on distribution of species in space and time, quantitatively if possible, from both physical collections and observational databases if possible. (Collections have the advantage of being verifiable, while observational databases tend to have vastly larger numbers of records.) Other information (such as behaviour, condition, etc.) will be in much less demand but critical for specialists. Typically, conservationists are either interested in species protection, hence in the distribution of a species or set of species within a specified region, or in site protection, hence the list of species for a given area (whether a geographic area or a nature reserve, etc) will be desired. This means that the key fields for sharing data are (1) date of observation, (2) coordinates, and (3) taxon name. Although there may be taxonomic difficulties and differences in coordinate systems, this limited number of fields can be readily shared in networks across databases or queried simultaneously, pooling information from multiple sources and increasing its power.

Networks have developed so rapidly that the options and opportunities to use them are becoming confusing, especially since much information about them is in the "grey literature" of the Internet, rather than traditional professional publications. With this in mind, we present a review of the available types of information networks, and an example from each. Because this paper developed from a workshop on information sharing in Asia, our focus is on the use of the networks relevant for Asian conservationists.
Table 1: Some examples of online information services for conservation.

<table>
<thead>
<tr>
<th>Name</th>
<th>Web site</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBIF</td>
<td><a href="http://www.gbif.org">www.gbif.org</a></td>
<td>Global biodiversity facility, unit level records</td>
</tr>
<tr>
<td>Hotspots</td>
<td><a href="http://www.biodiversityhotspots.org">www.biodiversityhotspots.org</a></td>
<td>Global</td>
</tr>
<tr>
<td>IUCN SIS</td>
<td><a href="http://www.iucn.org/themes/ssc/our_work/sis.htm">www.iucn.org/themes/ssc/our_work/sis.htm</a></td>
<td>Global</td>
</tr>
<tr>
<td>IUCN Redlist</td>
<td><a href="http://www.iucnredlist.org/">www.iucnredlist.org/</a></td>
<td>Global</td>
</tr>
<tr>
<td>UNEP-WCMC</td>
<td><a href="http://www.unep-wcmc.org">www.unep-wcmc.org</a></td>
<td>Global, multiple databases and information types</td>
</tr>
<tr>
<td>BioCASE</td>
<td><a href="http://www.biocase.org">www.biocase.org</a></td>
<td>European, includes metadata and unit level data</td>
</tr>
<tr>
<td>Natureserve</td>
<td><a href="http://www.natureserve.org">www.natureserve.org</a></td>
<td>Americas, relevant to Asia as a model</td>
</tr>
<tr>
<td>MRC Mekong</td>
<td><a href="http://www.mrcmekong.org">www.mrcmekong.org</a></td>
<td>Regional Special Interest Network</td>
</tr>
<tr>
<td>CSIS</td>
<td><a href="http://www.chinabiodiversity.com/search/english">www.chinabiodiversity.com/search/english</a></td>
<td>National</td>
</tr>
</tbody>
</table>

Asia has particular needs for development of, participation in and access to such networks for several reasons:

- Asia possesses several biodiversity hotspots (Conservation International 2006, Mittermeier et al. 1999; Myers et al. 2000; Mittermeier et al. 2004), yet much of the data on them reside in non-Asian countries;
- Asia is a vast continent and economic resources are usually quite limited. Practically speaking, it is difficult for conservationists to travel and meet for information sharing. On the other hand, Asia is relatively technically adept, with access to internet and email in even remote areas. Hence electronic networking helps the conservation community share information and expertise;
- Many Asian species and ecosystems of high conservation value are multinational (e.g. tigers, snow leopards, river basins, mountain ranges), making regional international coordination increasingly important in conservation management.

**TYPES OF INFORMATION NETWORKS**

Information networks are needed to meet these needs. These can be categorized as regional, special interest and global networks. Regional and special interest networks tend to be focused on particular management goals and can be complex. Global networks may be either specialized and somewhat complex, or generalized and fairly simple in structure.

Information networks may include news, expertise, searchable databases, and any other kind of material which can be put on web sites. We confine ourselves to those networks with
searchable databases in order to highlight what we may call *conservation bioinformatics* (paralleling the “Biodiversity Informatics” and "Ecological Informatics" mentioned above).

The newly emerging field of bioinformatics includes research, development, or application of computational tools and approaches for expanding the use of biological data, including those to acquire, store, organize, archive, analyze, or visualize such data (NIH working definition of bioinformatics and computational biology, BISTIC Definition Committee, 2000 ). Although this concept is most commonly applied to biochemical and genetic research, it seems equally useful within the context of biodiversity, conservation and ecology.

**Examples of each network type**

*Special Interest networks:* A Special Interest Network (SIN) is a set of network sites ("nodes") that collaborate to provide a complete range of information activities on a particular topic (Green 1994). SINS are emerging as an important new paradigm for large-scale collaboration on the Internet. Coordination is achieved through logical design, automation, mirroring, standards, and quality control. To be successful, SINs should strive to provide reliable and authoritative information services, to encourage participation, and to accommodate growth (Green 1994). Examples of SINS in conservation include enforcement and wildlife trafficking, endangered species databases, expertise databases, etc.

An example of a SIN: Mekong River Commission Information System. Data networking has been used effectively in the Lower Mekong River Basin (LMB) for ecosystem management at a large catchment scale. The MRC-Information System (MRC-IS) primarily supports the implementation of the ‘Agreement on the Cooperation for the Sustainable Development of the Mekong River Basin,’ signed by the LMB countries (Cambodia, Lao PDR, Thailand and Vietnam) in 1995 (Mekong River Commission, 2005).

The goals of this data networking are: 1) to provide integrated databases of all relevant data and information required for the implementation of the Mekong Agreement; 2) to develop and support models to enable analysis and generate information and knowledge for decision making within the Lower Mekong Basin; 3) to develop, support and promote institutional and technical mechanisms for data and information exchange and sharing; and 4) to enhance capacity in the fields of data and information management.

The data types involved are diverse and derive from a number of resource types including GIS spatial data, monitoring of hydro-meteorological data through space and time, socio-economic statistics, remote sensed images, and a document database (Fig. 1). The only practical way to integrate such diverse data types is geo-spatially, and so the overall network and sharing architecture is GIS-based.

The web-based MRC-IS (see [www.mrcmekong.org](http://www.mrcmekong.org) and [www.mekonginfo.org](http://www.mekonginfo.org)) is donor funded, but oriented towards local needs. In this region, use of the Internet is the only practical way to share data. The MRC-IS is also highly GIS-oriented for data management, and includes extensive metadata. Hence, the system design requires a center with high technological capacity, which in turn is oriented to enable users without high technological capacity to access an array of information, data and documentation.
**Regional and National Networks:** Some countries and sets of countries have begun to link databases that contain biodiversity information of similar types or to develop comprehensive stand-alone biodiversity databases. Conabio (Comisión nacional para el conocimiento y uso de la biodiversidad 2006), BioGIS (Hebrew University of Jerusalem 2003), and BioCASE (BioCASE Secretariat, 2006) are all examples of this kind of networking. In most cases, natural history specimen and observational databases from multiple sources are either joined by an XML query system or merged in a central repository for online queries. Cutting-edge programming technology has been employed to make multiple-database queries feasible, and much of the technology is now available to the public. The China Species Information System (CSIS) provides an excellent example of a detailed national-level database. These databases are mainly focused on providing the user with the fundamental information of “what occurs where at what time,” e.g. taxon name, location and date for records in either collections or observations.

**Example: BioCASE: The Biodiversity Collection Information Access Service for Europe (BioCASE Secretariat, 2006)** is a regional network for access to biological collections and observational databases in Europe and Israel. The value of this system for Asia (apart from Israel) is that many specimens collected in Asia reside in European collections, providing information on what may have existed in Asia in the past (information repatriation).

BioCASE contains two parts: a metadata database about the collections and databases; and a simultaneous query system for digitized unit level records (specimens and observations) where these have been contributed to the system. According to its self-description, BioCASE is a transnational network of biological collections of all kinds. BioCASE enables widespread unified access to distributed and heterogeneous European collection and observational databases using open-source, system-independent software and open data standards and protocols (BioCASE Secretariat, 2006). It should be noted that BioCASE is now the Global Biodiversity Information Facility (GBIF) network for Europe (see GBIF, below), and thus
became nested within a global biodiversity network for unit level data, but remains unique in its metadata provision about collections. The BioCASE unit-level portal was subsequently restricted to European biodiversity records, thus distinguishing it from the GBIF global portal. BioCASE portals can easily be further regionalised and thematically restricted, so for example a botanical portal for German biodiversity has been developed and the same technology will now be used by the French national GBIF node.

Example: CSIS: The China Species Information Service (www.chinabiodiversity.com) is an example of a detailed and comprehensive national biodiversity database. CSIS was started in 1996 by the Institute of Zoology, Chinese Academy of Science (IOZ/CAS) as the China Endangered Species Information System (CESIS). During 1998-2004, it was expanded into the CSIS by the Biodiversity Working Group (BWG) of the China Council of International Cooperation for Environment and Development (CCICED). It is now maintained by the China Program of the Wildlife Conservation Society and the IOZ/CAS. This database currently consists of six main elements, each of which can be searched on-line. These are:

- **China Species Red List** contains a comprehensive and complete data set including most major taxa. The number of records includes 2,225 species of invertebrates (prawns, corals, mollusks, spiders, jellyfish, scorpions, crabs, butterflies, beetles), 2,663 species of vertebrates (709 fish species and all Chinese terrestrial vertebrates), and 4,364 species of plant (all gymnosperms, orchids, rhododendrons, Gesneriaceae and *Acer*, and some species of other groups). All these species have been evaluated following IUCN Red List Criteria by over 100 taxonomists (see Wang & Xie 2004). Listed information includes: taxon, synonym name, endangered category (including China Red List, IUCN Red List, CITES appendix, rank in Chinese Red Data Book, state protected species list), Chinese name, English name, distribution, habitat, threats, protective measures taken, recommendations, GIS point map, and photographic images (if available);

- **Alien Invasive Species in China** contains information on over 450 alien invasive species from most major taxa. Information includes: taxon, synonym name, Chinese name, English name, country or region of origin, distribution in China, path of spread, impact, control measures, GIS point map, and photographic images (if available);

- **Protected Areas in China** contains information on 1971 nature reserves. Information includes: locality of China’s protected areas, their management authority, size, altitude, year established, administrative category, species list for some nature reserves, conservation targets, introduction, ecotourism, contacts (if available), and photographic images (if available; see Xie et al. 2004; Wang & Xie 2005);

- **Species Identification** contains search tools for identifying amphibians, reptiles, birds and mammals. This database allows users to search by visible information on species identification. Instead of following a taxonomic-based identification manual, a procedure that is even difficult for taxonomists to use, the electronic identification guide is based on easily-recognized characteristics such as color, size, ratio of tail to body length, and distribution range;

- **Expert information** giving contact information for specialists on Chinese flora and fauna; and

- **Bibliographic information.**

The CSIS is increasingly being used for biodiversity analyses of Chinese flora and fauna and to support conservation tools for conservation biologists and policy-makers in China (such as
the China Red List and China’s Protected Areas cited above). It was a critical tool in the production of *The Birds of China* (MacKinnon & Phillipps, 2000) and the upcoming *The Mammals of China* (Smith & Xie, in production).

**Global biodiversity networks:** By far, the networks of most general interest and utility are global networks. The focus of global biodiversity networks is to provide compilations of data, as well as broad overviews of global biodiversity. The range of potential users is broader than for special interest networks, and often for regional ones; they include academic researchers, governmental agencies, collection curators, educators, and economic enterprises as well as conservationists. These data networks are partly a natural outgrowth of the basic kind of data kept by most biologists: what organism is observed, when, and where. They also have developed in direct response to Article 17 of the Convention on Biological Diversity, on Exchange of Information. Here, we present two examples, one fairly focused and one that is broadly inclusive.

**Example: The IUCN/SSC Species Information Service, A global network for conservation.**
The Species Information Service (SIS; IUCN 2006) is being developed by the IUCN Species Survival Commission specifically as a system to support conservation of the world’s biodiversity resources. Through its network of 8000 volunteer species-conservation experts around the world (approximately 1500 in Asia), the SSC holds the most comprehensive body of current data and information in existence on the status and distribution of those plant and animal species of special conservation concern. Not only is SSC uniquely positioned to capture, manage and analyze a massive resource of biodiversity data and information, but it is also well-positioned to ensure that the emergent information supports biodiversity conservation decision-making processes. SSC is one of six expert Commissions within IUCN (the World Conservation Union), and as such has ready access to the IUCN membership of 77 sovereign states, 112 government agencies, and over 800 non-governmental conservation organizations from throughout the world. Through IUCN, SSC also has privileged access to a number of bi- and multi-lateral organizations (e.g., the World Bank) and to international treaties that relate to conservation and sustainable development.

The Species Information Service (SIS) is designed to provide the tools needed for SSC network members to manage data in consistent formats, exchange data and information efficiently, and pool data and information to produce high-quality, sophisticated biodiversity analyses. In essence, it functions in much the same way as a special-interest network as described above, but is global in extent. SIS enables SSC to produce biodiversity information products in cooperation with other conservation organizations, and to respond more readily to requests for specific information products needed for decision-making processes.

SIS is unique among global biodiversity networks in its focus on using current data to support conservation of biodiversity. No other biodiversity information system includes a current source of biodiversity information that covers the whole world and accommodates such comprehensive information for a range of taxa. SIS is designed to capitalize on existing species biology and bioinformatics knowledge found throughout the SSC network, and on existing partnerships with other conservation organizations and networks that hold complementary data and information. To facilitate the incorporation of species data into SIS, two independent software applications have been developed: the Data Entry Module (DEM) and the Red List Database (an extension of the ongoing IUCN/SSC Red List of Threatened Species). The DEM was used to construct the Global Amphibian Assessment (IUCN *et al.* 2004), one of the most thorough data compilations to date about a taxon with accompanying threatened species status. Various other assessments are also underway (Global Mammal Assessment; Global Marine Species Assessment). These assessments have been greatly facilitated by a number of regional workshops held throughout Asia.
To better address biodiversity protection and global change analysis requirements, the SSC is developing an enhanced version of SIS that will be web-enabled as a comprehensive biodiversity information resource. This version will provide current, spatially explicit biodiversity information to support scientific discovery, natural resource management, and policy formulation, as well as to enable the SSC to improve measurements and monitor changes in biodiversity over time. Analyses can be produced at a variety of scales (local to global), ensuring their relevance to those organizations and government agencies responsible for conservation and development. The project team for this new development activity includes Conservation International, NatureServe and BirdLife International working under the coordination of SSC/IUCN.

A major objective is to enable the SSC network, through its sub-networks (Specialist Groups) to gather data and disperse information in a standard way. It will provide the technological infrastructure necessary to fulfil the data management and flow needs of the IUCN Species Survival Commission and its network of volunteers. Asian specialists will be able to reap immediate benefits from the two-way flow of information that is built into the SIS structure.

Example: The Global Biodiversity Information Facility (GBIF): A global network for biodiversity data compilation. GBIF (GBIF Secretariat 2006), is broad in scope and currently more limited in information type than the SIS, but potentially includes vastly more records and linkages to a vast array of information types including molecular, ecological, ecosystem and digital library resources. GBIF’s data portal, which is already available online as a prototype, provides integrated primary record level (label or observational) data on all species of living organisms. It does this by providing software and web services (using extensible markup language, or XML) that enable integration of data drawn from multiple resources that are distributed around the world. Unlike the SIS, which is providing common data-entry software (DEM) and a major, unified database, GBIF provides software that allows any data provider using any platform and any database engine to serve their data through the GBIF system and have it integrated with such data from many other data providers.

The GBIF mission statement is "to make the world's primary data on biodiversity freely and universally available via the Internet" (www.gbif.net). GBIF currently focuses on global species databases (GSDs) and checklist data to generate a complete Electronic Catalogue of Named Organisms, and on the mobilization of primary species occurrence data ("unit-level data", i.e. specimen or observation records). The basic conservation biology challenge that GBIF addresses is the disparity between the biodiversity hotspots and the location of data about them (Fig 2).

GBIF is the most ambitious of all the projects described here regarding integration of biodiversity data. Like BioCASE, it is designed to query multiple databases via a unified query system. It works via both national focal points and networks (for example, in partnership with the previously-mentioned BioCASE, IUCN, and Conabio). Currently, 47 countries and 32 international organizations are members in GBIF, and the membership steadily grows, as does the data content (as of 15 April 2006, it contained 91.5 million occurrence records served by 159 data providers in 35 countries available for access via the GBIF network).

Example: An Information Center with multiple databases: the World Conservation Monitoring Centre. A somewhat different approach is used by the United Nations Environment Programme-World Conservation Monitoring Centre (UNEP-WCMC) in Cambridge, England. UNEP-WCMC is the biodiversity assessment and policy implementation arm of UNEP (UNEP-WCMC 2006). Founded by IUCN, WWF, and UNEP, WCMC has pioneered much of the effort in establishment of online information networks, serving a variety of conservation organization, most notably IUCN and now UNEP. Rather than providing a network of diverse databases accessible by a single query system, UNEP-
WCMC provides a web site with access to multiple databases of global biodiversity information. Its species databases include the "Threatened Plants Database," "Tree Conservation Database," “Threatened Animals of the World,” and “CITES Trade Database”, supplemented by information on protected areas (World Database on Protected Areas), forests, and marine systems (especially corals and mangroves). Alongside the access to information through text-based websites, UNEP-WCMC makes extensive use of interactive map web sites and is increasingly partnering with organisations by XML interoperable links. Since WCMC is an active center of development for information systems, this list is necessarily partial and in a state of regular expansion.

**Figure 2:** The major natural history museums and herbaria (blue dots) that hold primary information about the species that inhabit the identified biodiversity hotspots of the planet are not located close to those hotspots (red dots). However, the data that accompany the specimens in those major collections, and other biodiversity data, can be delivered to the hotspots for use “on the ground” via the Internet if they are digitised. GBIF is promoting this digitization and delivery on a global basis

- biodiversity hotspot
- holder of large amounts of biodiversity data

**DISCUSSION**

These different types of networks have some commonalities. Conceptually, most are based on multinational cooperation of teams of biologists and technical persons. They use cutting edge technology which is under rapid development. This new technology influences conceptualization of what the networks can do and accomplish, and both are evolving. In all the systems presented, the Internet is central to information provision, eliminating distance as a barrier. In many cases, different components of the network are located in different countries.
The content of the networks is mainly of two types, primary "raw" data or secondary "interpretive" or "expert" data. Both types are used in the more specialized or management-oriented networks. Simpler primary data are emphasized in geographically broader, more general data provision. It is important to understand that the linkages that are envisaged for the future, in addition to the analytical applications of today, are dependent on two fundamental aspects of biodiversity data and information: the “key” nature of scientific names of organisms and the unifying nature of geospatial referencing within the data themselves.

Biodiversity information networks generally are moving toward linking to each other through various gateways, with the wider vision of simple public access to all available biodiversity information. Indeed, GBIF and WCMC, with Cardiff University and Microsoft Research, are developing an integrated toolkit that will enable users everywhere, who access the broader system through either GBIF or WCMC (or indeed one or more of the regional networks) to utilize web-based tools for biodiversity analysis. Based on a workflow model, the Ark2010 project aims to have this open-source toolkit available for use sometime in 2007.

Perhaps more importantly than common tools development is the general consensus in biodiversity informatics that open data access is not only desirable but increasingly a necessity to facilitate conservation efforts. To this end, GBIF, IUCN and other networks and organisations have ascribed to the principles of the Conservation Commons (see http://www.conservationcommons.org/), and GBIF and the COP of the Convention on Biological Diversity (among other organizations) have issued statements on open access (see http://www.gbif.org/Stories/STORY1138028174) and decisions (UNEP/CBD/COP/8/L.5).

Another very important development is the increased effort put into standardisation of data and information systems within the community, spearheaded by the “Taxonomic Databases Working Group”, traditionally named but now extending far beyond taxonomy. For example, all standards used for data encoding and transmission used by GBIF are TDWG standards (TDWG 2006).

Though the trend for the foreseeable future is toward distributed systems, it often is effective to centralize the personnel who are working on information systems, as demonstrated by the Mekong River Commission, GBIF, CSIS and UNEP-WCMC. A critical mass of information specialists and programmers is usually needed to initiate and run complex information systems even if the components become dispersed later.

Note that all these features tend to make data networks inherently somewhat anarchic. The lack of centralization, the heavy use of the Internet, and the ability to construct sophisticated databases on limited resources all render centralized “control” difficult. Thus, centralized top-down authorities may find the approach threatening. Reliance on trust and professionalism without much possibility of enforcement of standardized use make these endeavors inherently appealing to academics and the current generation of Internet users, and thus promote the growth of information content and expansion of the systems. However, political resistance could be anticipated as this growth occurs.

In balance against this is the trend towards connection of information networks, providing universal access through various gateways. Agreements are worked out between networks for terms of cooperation, usually involving mutual information sharing and accreditation of data sources. Far from being wildcat data mining, the conduct of most researchers involved in these endeavors tends toward consensus and mutual respect. We find this encouraging. In the absence of external regulation, the information communities appear to agree on self regulation for the mutual benefit of all. We can only hope this trend continues.
ACKNOWLEDGEMENTS

We wish to thank Dr. Phillip Fox of UNEP-WCMC and Dr. Eliezer Frankenberg of the Israel Nature and Parks Authority for reading and commenting on the manuscript. This paper is derived from the special workshop on Data Networking for Conservation Biology in Asia, part of the Society for Conservation Biology – Asia Section Symposium on "Biodiversity Conservation in Asia: Current Status and Future Perspectives" held in Katmandu, Nepal during 17–20 November 2005. We thank the participants in the workshop for their valuable input and support during the conference.

REFERENCES


Green, D.G. 1994. A web of SINs - the nature and organization of Special Interest Networks. [Note: this is an incomplete reference]


