



The Data-Sharing Toolbox

By Falk Huettmann, Ph.D., and Zachary Meyers



Courtesy of Falk Huettmann

Falk Huettmann, Ph.D., is Associate Professor at the Institute of Arctic Biology in Alaska.



Courtesy of Zachary Meyers

Zachary Meyers is a master's student at the University of Alaska, Fairbanks.

The array of data-sharing technologies today can be totally overwhelming. Yet the need for accessing the latest research and news about wildlife and biodiversity is more critical than ever. What follows is our analysis of some of the most useful tools available to wildlife biologists, managers, and policymakers who are on the hunt for relevant data.

Open Sesame

The critical need to share and manage biodiversity knowledge gained global attention in the early 90s which led to the formation of the [Convention on Biological Diversity](#). This treaty was designed to encourage the gathering, management, and open sharing of biodiversity data. Over time, signatories to the treaty saw the need for a mechanism that would allow access to a comprehensive inventory of biodiversity data, with information on subjects that could be as broad as the migration patterns of bald eagles (*Haliaeetus leucocephalus*) or as localized as population shifts of the American pika (*Ochotona princeps*) on a particular mountaintop.

Then, in 2001, the [Open Society Institute](#) launched the [Budapest Open Access Initiative](#)—the first and foremost international statement calling for open access to all academic research and information. That same year, a working group of the Organization for Economic Cooperation and Development requested the formation of a system that would make primary biodiversity data available and accessible anywhere in the world via the Internet. To that end, a partnership of countries and international organizations launched the [Global Biodiversity Information Facility](#) (GBIF), a free and openly accessible infrastructure that has grown to include 50 member countries and 40 international organizations. To date, GBIF provides access to more than 181 million primary biodiversity records residing in over 8,000 datasets stored in approximately 294 institutions around the world.

A Guide to Databases

It's the age of the database. Numerous databases, web portals, and datasets exist with extensive amounts of data and information. Researchers must access multiple databases to get comprehen-

sive, up-to-date data on their specific area of study. Here, we provide a sampling of the databases and datasets designed exclusively to provide comprehensive biodiversity data.

- The [Mammal Network Information System](#) (MaNIS) was the first of four databases created to provide data on vertebrate specimens stored in museums across the world. Established in 2001 and funded by the National Science Foundation, MaNIS is based on the [SpeciesAnalyst](#) prototype—a software tool developed by the University of Kansas Natural History Museum in 1999.

Currently 67 institutions are registered to supply data to MaNIS users. Each of these institutions holds a vast number of specimens. The Texas Cooperative Wildlife Collection, for example, houses approximately 59,000 mammal specimens while UCLA's Ronald R. Dickey Collection of Birds and Mammals consists of more than 50,000 skeletons and skins of birds and mammals from the U.S., Canada, Mexico, Central America, and the Pacific Islands. MaNIS provides information on mammal collections that can range from the Royal Ontario Museum's first record of the smoky sheath-tailed bat (*Cyttarops alecto*) to the University of Michigan's 1935 Field Notes on the southern flying squirrel (*Glaucomys volans*). A researcher studying the behavior of the common pipistrelle (*Pipistrellus pipistrellus*), for example, can send a request to one, several, or all of the 67 participating institutions, which will then return the search results with extensive amounts of data stored on each institution's server.

- [HerpNet](#), the second of the four vertebrate databases to be established, provides data on herpetological collections across the world. Similar to MaNIS, HerpNet includes geo-referencing tools that can help researchers identify exact locations of their species of interest.
- The [Ornithological Information System](#) (ORNIS), established in 2004, provides observational and voucher-based data for bird species around the world. ORNIS contains 36 million observational records and 3.9 million specimen records spread across 29 collections. Unlike the other three databases, ORNIS also provides sound recordings.



- **FishNet2**, also established in 2004 and funded by an NSF grant, provides extensive data on 20 million fish specimens housed in 29 institutions worldwide.
- **Encyclopedia of Life (EOL)**, launched in 2007, was created to organize and improve access to what its creators—scientists across the world—refer to as “virtually all information about life present on Earth.” EOL is a website of websites that serve as entry points into information on the approximately 1.8 million known species of plants and animals. So, for example, if a biologist were to run a search on a cougar (*Puma concolor*), the search would return detailed information on the description of the animal—its morphology, reproduction and life history traits, and behavior—along with the ecology and distribution of the species as well as its current conservation status (*Puma concolor*, EOL). The EOL staff and content providers include scientists and non-scientists around the world.
- **Data Basin** is a web tool that connects scientists, practitioners, and citizens with datasets, tools, and information on conservation data. In addition to providing access to maps and charts, Data Basin stores biological, physical, and socio-economic datasets, which a user can explore by geography, keyword, or topics. Data Basin users can also build professional networks by setting up “expertise profiles” that the rest of the user community can access.
- The **Catalogue of Life** is an international partnership composed of the **Integrated Taxonomic Information System (ITIS)** and **Species 2000**, both of which provide taxonomic data on the world’s plant and animal species. The goal of the Catalogue of Life is to serve as a comprehensive reference list to the scientific and common names of all known species on Earth. Every year, the catalogue will create an annual checklist; the 2008 version contains information on about 1.3 million species. The Catalogue of Life also serves as the taxonomic backbone for EOL.
- **NatureServe** is a non-profit organization that represents an international network of biological inventories that operate across the U.S., Canada, and Latin America. The organization manages detailed local information on plant and animal species and ecosystems, providing conservation status assessments for each species at global, national, and state scales. It also develops data management tools such as GIS mapping to address conservation needs. Additionally, the **NatureServe Explorer**



Credit: NatureServe

The homepage of NatureServe’s online encyclopedia allows users to search through data on more than 70,000 plants, animals, and ecosystems in North America. The database provides conservation status assessments for each species on global, national, and local levels.

website contains data on 70,000 plants, animals, and habitat types in North America.

The Missing Pieces: Metadata

A wealth of data may remain hidden if it can’t be easily searched and found. The solution to this problem is metadata, or data about data. Metadata is somewhat analogous to an old-fashioned library index card, which lists the author, title, publication date, and shelf location of a book. Scientific metadata might include the name of the researcher, keywords about the species or phenomenon being addressed, and the date, location, and purpose of the research. Though such information is vital to conducting speedy, accurate searches, many researchers still aren’t in the habit of including it with their research findings.

Prior to the advent of the Internet Age, wildlife research suffered from a lack of standardization in the collection, analysis, and publication of data. To some extent it still does. In the last decade, however, advances in information technology have caused something of a revolution in the way biodiversity information is created, maintained, distributed, and used (Bisby 2000).

Standardization helps ensure that data are characterized in the same manner and can allow researchers to integrate data from multiple sources. Today several initiatives are under way to address the urgent need for



standardization of data and metadata. Among them:

- The **Dublin Core Metadata Initiative** (DCMI) is an organization that works almost exclusively on developing and maintaining an international standard for describing online data and information.
- **Darwin Core**, based upon the ideas and principles of the Dublin Core, is a biodiversity informatics data standard that contains an extensive vocabulary of terms to facilitate the discovery, retrieval, and integration of information about an organism.
- The **Distributed Generic Information Retrieval** (DiGIR) protocol is an implementation of the Darwin Core that allows contributing researchers and biologists to link their data to web portals, while staying in full control of their data in general along with the extent of data sharing and the level of data delivery. DiGIR also enables users to connect and query computers via a web portal. MaNIS, ORNIS, FishNet, HerpNet, and a host of other web portals all use DiGIR.
- The National Biological Information Infrastructure (NBII) has developed what it calls the NBII biological metadata standard, which outlines how to describe biological data and information. That

information is then added to the **NBII Metadata Clearinghouse**, which serves as an online and searchable card catalog with descriptions of numerous different biological datasets from multiple agencies and organizations (See Tools and Technology, page 62).

- The **Federal Geographic Data Committee** develops procedures and assists in the implementation of a distributed discovery mechanism for national digital geospatial data for use in such applications as geographic information systems or GIS.
- The **Data Documentation Initiative** is an effort to develop a standard to describe social science data.

Pride of Ownership

Although in an ideal world all scientific data would be openly available to all researchers, this isn't always the case. Users attempt to make their data easily available, while protecting sensitive information (such as the location of the last population of an endangered species). Today there are several web tools that allow for both sharing and control. For example, **Creative Commons**, an organization launched in 2001, provides free licenses and other legal tools that define how information can be used by others. It allows data owners to determine how others can share, remix, or reuse their data. Similarly GBIF, a global information infrastructure that facilitates free and open access to the world's biodiversity data, guarantees that its data publishers will retain control of their own data. In general, most databases today are designed in such a way that data holders and publishers retain ownership and management of their data, while allowing access to the data that they choose to make publicly available.

Clearly the revolution in data sharing and management is here and will continue to evolve. Whether using web portals, GIS systems, search engines, clearinghouses, or other tools to record, manage, share, and find data, today's wildlife professionals have multiple data-sharing resources at their fingertips, which ultimately will help advance the cause of wildlife conservation. ■

Additional reporting by Divya Abhat, science writer for The Wildlife Professional.


Western EcoSystems Technology, Inc.
Specializing in the design, analysis, and conduct of ecological field surveys



2003 Central Avenue
Cheyenne, Wyoming 82001 307.634.1756
www.west-inc.com

WEST, Inc. offers clients a unique combination of field ecology and statistics to help solve natural resource problems

- Biometrics/Statistics Expertise
- Resource Selection Research
- Threatened and Endangered Species Work
- Wind Power-Site Placement and Monitoring Studies
- National Environmental Policy Act Work
- Wetlands & Water Quality Studies
- Radar Studies
- Large Mammal Research
- Capture-Recapture Studies
- Avian and Bat Studies

See this article online at www.wildlife.org for additional literature on biodiversity data sharing and management.