

- AND O. D. VEPRINTSEVA. 1996. Natural sound archives: Guidance for recordists and request for cooperation. Pages 474–486 in *Ecology and Evolution of Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, Eds.). Cornell University Press, Ithaca, New York.
- NELSON, D. A., AND S. L. L. GAUNT. 1997. The Borror Laboratory of Bioacoustics (BLB) and the Bioacoustics Research Group at The Ohio State University. *Bioacoustics* 8: 281–286.
- NELSON, D. A., S. L. L. GAUNT, C. L. BRONSON, AND T. J. KLOTH, JR. 2001. Database design for an archive of animal sounds. *IEEE Engineering in Medicine and Biology* 20:76–80.

---

*The Auk* 122(3):987–990, 2005  
 © The American Ornithologists' Union, 2005.  
 Printed in USA.

**Free and Open Access to Bird Specimen Data: Why?**—Ornithology is in a unique position in systematics. Birds are the only major taxon for which more than 99% of species taxa at every point on the surface of the Earth are likely to be known to science (Mayr and Vuilleumier 1983, Peterson 1998). Scientific collections of birds document the distribution and diversity of more than 10,000 species worldwide. Although even these collections are in need of augmentation and improvement (Remsen 1995, Winker 1996, Peterson et al. 1998), data associated with existing specimens constitute a rich source of information about avian distribution and diversity. This resource could serve as the basis for many exciting analyses and insights into the natural history, ecology, systematics, and conservation of birds (Remsen 1995), and as a guide and motivation for further improvement of the specimen basis and information resources.

The need for more efficient access to ornithological data, however, is great. Systematic efforts to document and study avian diversity rely on the specimen record as a critical guide. Biodiversity conservation efforts depend heavily on avian information, as bird distributions can inform conservation planning and prioritization much more completely than other, less well-known taxa. Numerous other applications in natural history, biogeography, ecology, natural

resources management, and even public health also draw insights from avian data (Rappole et al. 2000). This situation thus calls for an efficient system serving accurate ornithological information broadly, both to meet such varied needs and to demonstrate the critical importance of the resource that underlies them.

Presently, such a system does not exist. For example, recent efforts to assemble a list of all specimens of Red Junglefowl (*Gallus gallus*) in natural-history museums in North America and Europe took six and a half months of letter-writing and e-mailing to result in a list of 752 specimens (Peterson and Brisbin 1998). Similarly, efforts to assemble large-scale data sets on migratory bird breeding and wintering areas, necessary for modeling the future distribution of West Nile Virus in North America, were stymied by inefficient access to information and took many months of effort and unnecessary tricks of data manipulation (Peterson et al. 2003).

The technology for such a biodiversity information system nonetheless exists; it was, in fact, developed on the basis of avian data sets, with funding from the National Science Foundation. Subsequently, several efforts have begun assembling such systems across many taxa (see Appendix). Most exciting is that developers of these systems have collaborated to develop a next-generation technology that will meld all these regional efforts into a single, global biodiversity information system—the technology, termed “DiGIR” (distributed generic information retrieval), has won broad acceptance and has been incorporated into many efforts.

Ornithology, with its large quantities of high-quality information regarding an important indicator taxon, has the opportunity to lead this new world of biodiversity informatics. Several other taxonomic communities have already advanced in integrating their data resources via the internet (for examples, see Appendix), and several institutions have already ventured their ornithological data resources in a prototype internet-based distributed system (The Species Analyst, now superseded by ORNIS). Nevertheless, many computerized ornithological data sets remain either unavailable over the internet or available, but not integrated with data sets from other institutions.

*Free and open access and data value.*—Biodiversity information has traditionally been concentrated in Europe and North America,

even though biodiversity is focused in tropical and subtropical regions. This contrast results from the complexities of the history of scientific exploration, economics, and educational and scientific opportunities. Like biodiversity itself, access to information about biodiversity is unbalanced.

Modern internet technologies make feasible a system in which information resources can be accessed by anyone, anywhere on Earth. The internet provides a medium of information flow that is limited only by internet access, a barrier that is rapidly disappearing over much of the planet. Hence, the regional imbalances that characterize the current situation can be largely alleviated.

The key point of most debates on the subject of free and open access has been the value of specimen data (Graves 2000). Museum curators know that the information associated with the specimens they curate is valuable, and for that reason they have often guarded such information carefully—the limited budgets at most collections, many of which are in serious financial situations (e.g. recent problems at the Academy of Natural Sciences, Philadelphia), demand that any resource be used wisely. Moreover, resources dedicated to computerization and broad data provision may occur at the expense of specimen care and building the collection itself. However, “valuable” data that are not used yield nothing to the owners or curators of those data.

By contrast, data that are used increase markedly in value. Biodiversity information is too often derived from secondary sources (range maps, field guides, etc.), which both reduces data quality and denies credit to those institutions that house the primary data (often natural-history museums). A system with free and open access to data, however, permits users to access the primary, vouchered information as close to its source as possible. Similar to the marketing strategies of Netscape and Adobe Acrobat, in which providing free and open access is instrumental in building a market share and making a product, such access is key to establishing natural-history museum collections as the premier source of information about biodiversity.

In this sense, the value of data does not decline, but rather increases, as a result of free and open access. That is, as primary ornithological data from specimens become the primary source of information on the distribution of birds, those

data gain value. Furthermore, open access to specimen data results in feedback that leads to higher quality, again increasing the value. By contrast, data for which access is restricted do not benefit to the same extent from analysis, scrutiny, feedback, and interest.

*Distributed, not centralized.*—A key feature of the information systems under discussion is their distributed nature. Distributed databases may be scattered across regions and countries, but are integrated via the internet. This structure offers distinct advantages: (1) data remain at the owner institution and are usually not centralized; (2) data served can be updated as often as desired, keeping information up-to-the-minute; (3) data ownership is never in question; (4) owner institutions can restrict or limit access as desired (e.g. to limit precision of data regarding distributions of endangered species, to protect rights of investigators regarding publication of works in progress, etc.); and (5) the collaborative nature of the effort is emphasized. Hence, although it required several years of dedicated activity to develop and distribute, this “architecture” makes the idea of providing free and open access to information much more palatable in a number of ways.

*Value added.*—Serving ornithological information is not a one-way interaction, not just a service to the broader community. Rather, uniting data resources into a single pool allows for several ways of adding value to the primary data. First and foremost, georeferencing locality information becomes much more feasible—because of the redundant nature of localities (specimens from single localities scattered across multiple collections, efficiency of georeferencing work on more densely collected landscapes), such an effort on a collection-by-collection basis is very inefficient. The success of efforts for georeferencing mammal specimen data (Stein and Wieczorek 2004, Wieczorek et al. 2004) is an excellent example. Several additional possibilities—use of ecological niche modeling to detect identification errors, standardization of taxonomic information, and use of collector itineraries to detect date–locality errors—are being developed. All these improvements to data can be repatriated to the owner institutions to improve the base quality of their data sets and information content of the specimens.

*Funding potential of community efforts.*—A particular advantage of community collaborations is

their excellent potential to leverage funding. The appeal of funding an effort in which all institutions in a community participate is much greater than that of funding an initiative that is based at a single institution. Clear evidence of this potential is the success that several taxonomic groups have had in getting funding for community efforts to integrate data: ichthyology, funded by the National Science Foundation and the Office of Naval Research; mammalogy, funded by the National Science Foundation; and herpetology, funded by the National Science Foundation—summing to more than \$4.5 million in new funding for informatics efforts in scientific collections. These resources would likely not exist without their community basis.

*ORNIS and the future.*—A fully integrated ornithological information infrastructure has enormous potential, and has now been funded by the National Science Foundation. Approximately  $4\text{--}5 \times 10^6$  bird specimens are held in North American museums, and ~80% of those specimens have been committed to participation in ORNIS. Perhaps yet another  $4 \times 10^6$  bird specimens are held in European museums, and an unknown quantity are held in museums elsewhere in the world ( $2\text{--}3 \times 10^6$  more?). Hence, a rough estimate is that on the order of  $10\text{--}12 \times 10^6$  bird specimens exist worldwide. If this resource were fully computerized and integrated into a distributed “world museum” of ornithology, the resource would be enormously useful in a broad diversity of applications. Integrating specimen-based data with observational data is enriching the specimen-based information still more: a recent addition to the ORNIS network included  $15 \times 10^6$  observational records from several projects based at the Cornell Laboratory of Ornithology.

At present, much information about birds is drawn from secondary sources. Conservation organizations prepare secondary information resources (lists of endangered species, distributional summaries, etc.). Field guides synthesize information into range summaries and distribution maps. Other resources are assembled solely on the basis of observational information, which lacks vouchering and can be unreliable in some circumstances (Phillips 1986). These secondary resources are too often used as the basis for answering important questions about birds.

Why are specimen data—the ultimate “library of life” information resource for biodiversity—

not already the primary information resource for birds? The answer lies in the difficult and inefficient access that has characterized this resource. Simply, the data are not used because they are hard to access. As ornithology provides better and more efficient access to specimen data resources—via ORNIS and related solutions, and their descendants—the user base will grow. Only in this way can avian collections get the key recognition and support they deserve and need.—A. TOWNSEND PETERSON, *Natural History Museum and Biodiversity Research Center, University of Kansas, Lawrence, Kansas 66045, USA (e-mail: town@ku.edu)* and CARLA CICERO AND JOHN WIECZOREK, *Museum of Vertebrate Zoology, University of California, Berkeley, California 94720, USA.*

#### LITERATURE CITED

- GRAVES, G. R. 2000. Costs and benefits of web access to museum data. *Trends in Ecology and Evolution* 15:374.
- MAYR, E., AND F. VUILLEUMIER. 1983. New species of birds described from 1966 to 1975. *Journal für Ornithologie* 124:217–232.
- PETERSON, A. T. 1998. New species and new species limits in birds. *Auk* 115:555–558.
- PETERSON, A. T., AND I. L. BRISBIN. 1998. Genetic endangerment of wild Red Junglefowl *Gallus gallus*? *Bird Conservation International* 8: 387–394.
- PETERSON, A. T., A. G. NAVARRO-SIGÜENZA, AND H. BENÍTEZ-DÍAZ. 1998. The need for continued scientific collecting: A geographic analysis of Mexican bird specimens. *Ibis* 140:288–294.
- PETERSON, A. T., D. A. VIEGLAIS, AND J. ANDREASEN. 2003. Migratory birds as critical transport vectors for West Nile Virus in North America. *Vector Borne and Zoonotic Diseases* 3:39–50.
- PHILLIPS, A. R. 1986. *The Known Birds of North and Middle America*. Part I. Privately published, Denver, Colorado.
- RAPPOLE, J., S. R. DERRICKSON, AND Z. HUBÁLEK. 2000. Migratory birds and spread of West Nile virus in the Western Hemisphere. *Emerging Infectious Diseases* 6:319–328.
- REMSEN, J. V., Jr. 1995. The importance of continued collecting of bird specimens to ornithology and bird conservation. *Bird Conservation International* 5:145–180.

- STEIN, B., AND J. WIECZOREK. 2004. Mammals of the World: MANIS as an example of data integration in a distributed network environment. *Biodiversity Informatics*, no. 4. [Online.] Available at [jbi.nhm.ku.edu/viewarticle.php?id=11](http://jbi.nhm.ku.edu/viewarticle.php?id=11).
- WIECZOREK, J., Q. GUO, AND R. J. HIJMANS. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 18: 745–767.
- WINKER, K. 1996. The crumbling infrastructure of biodiversity: The avian example. *Conservation Biology* 10:703–707.

#### APPENDIX

The following are websites for efforts to assemble biodiversity information systems: MaNIS ([elib.cs.berkeley.edu/manis](http://elib.cs.berkeley.edu/manis)); HerpNet ([www.herpnet.org](http://www.herpnet.org)); Global Biodiversity Information Facility ([www.gbif.net](http://www.gbif.net)); Red Mundial para la Información de la Biodiversidad ([www.conabio.gob.mx/remib/doctos/remib\\_esp.htm](http://www.conabio.gob.mx/remib/doctos/remib_esp.htm)); Virtual Australian Herbarium ([www.rbgsyd.gov.au/HISCOM/Virtualherb/virtualherbarium.html](http://www.rbgsyd.gov.au/HISCOM/Virtualherb/virtualherbarium.html)); SpeciesLink ([www.cria.org.br/projetos](http://www.cria.org.br/projetos)); European Natural History Specimen Information Network ([www.nhm.ac.uk/science/rco/enhsin/](http://www.nhm.ac.uk/science/rco/enhsin/)). For information on distributed generic information retrieval (DiGIR), go to [digir.sourceforge.net/](http://digir.sourceforge.net/).

For examples of taxonomic data resources on the internet, see [www.speciesanalyst.net/fishnet/](http://www.speciesanalyst.net/fishnet/) (ichthyology); [elib.cs.berkeley.edu/manis/](http://elib.cs.berkeley.edu/manis/) (mammalogy); [www.herpnet.org](http://www.herpnet.org) (herpetology). On efforts for georeferencing mammal specimen data, see [elib.cs.berkeley.edu/manis/](http://elib.cs.berkeley.edu/manis/). The ORNIS website is at [ornisnet.org](http://ornisnet.org).

both traditional and nontraditional uses of these shared research resources. For example, advances in analytical chemistry have enabled researchers to obtain data on heavy-metal contaminants and diets from a single feather. Future technological advances will increase nontraditional use of specimens, and two areas of rapid growth at present are in contaminant and stable-isotope studies. We address these developments and their implications for bird collections.

*Contaminants.*—Retrospective contaminant studies of the 1960s and 1970s premiered a new and important use of specimens. One of the first studies to use bird specimens in contaminant research documented a 10- to 20-fold increase in feather mercury among seed-eaters and raptors after the introduction of alkyl-mercury seed dressings (fungicides) in Europe in the 1940s (Berg et al. 1966). That research led to the banning of those seed treatments, and subsequent retrospective analyses using specimens confirmed the effect of alkyl-mercury fungicides by documenting the decline of mercury concentrations in feathers after the ban (Westermarck et al. 1975). Probably the best-known use of museum specimens in retrospective research documented eggshell thinning in raptors following the introduction of DDT in 1947 (Ratcliffe 1967, Hickey and Anderson 1968). These studies and others (see Kiff 2005) contributed to the eventual ban of DDT in many countries.

Researchers have documented high levels of contaminants in the biota of undeveloped regions, citing the global distribution of pollutants as the cause (Arctic Monitoring and Assessment Programme 1998). As global contaminant burdens increase, spatially and temporally distributed biological samples are needed to document changing contaminant levels. Archived avian specimens can document levels of heavy metals, because heavy metals bind to feather keratin at the time of growth (Crewther et al. 1965). Archived specimens were used to document increases in mercury pollution in several avian food webs (Appelquist et al. 1985; Thompson et al. 1992, 1993). Time series of archived seabirds were also used to document increases in feather mercury concentrations in two avian food webs over the past 100 years, which were correlated with anthropogenic inputs (Monteiro and Furness 1997, Thompson et al. 1998).

**Use of Bird Collections in Contaminant and Stable-isotope Studies.**—Preserved biological specimens are increasingly providing source material for research that is moving beyond traditional questions in collections-based studies. Technological advances are facilitating