

A call for the preservation of images, recordings, and other data in association with avian genetic samples, and the introduction of a solution: OMBIRDS

Authors: Szabo, Ildiko, Hurley, Grant, Cavaghan, Stephanie, and Irwin, Darren E.

Source: The Auk, 131(3) : 321-326

Published By: American Ornithological Society

URL: <https://doi.org/10.1642/AUK-13-237.1>

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



PERSPECTIVE

A call for the preservation of images, recordings, and other data in association with avian genetic samples, and the introduction of a solution: OMBIRDS

Ildiko Szabo,^{1*} Grant Hurley,¹ Stephanie Cavaghan,¹ and Darren E. Irwin^{1,2*}

¹ Cowan Tetrapod Collection, Beaty Biodiversity Museum, University of British Columbia, Vancouver, British Columbia, Canada

² Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

* Corresponding authors: Ildiko Szabo, ildiko@zoology.ubc.ca; Darren Irwin, irwin@zoology.ubc.ca

Received March 10, 2014; Accepted March 15, 2014; Published May 21, 2014

ABSTRACT

Much current and historical research in ornithology employs catch-and-release methods, resulting in a variety of data and materials from birds for which whole-body specimens have not been collected. Often, a genetic specimen (e.g., blood or feathers) is collected along with “media specimens” such as images and/or sound recordings, providing a rich source of research material as well as an opportunity to use each type of specimen as a source of validation of the other. Despite the abundance of these datasets and their potential use in future research, the preservation of such data and associated materials is currently a task that each researcher must confront individually, which results in the loss of these research materials over time. To promote the long-term utility of information collected from the thousands of birds that are captured and released each year, we present a protocol and database template (OMBIRDS; the Online Museum of Bird Images, Recordings, and DNA Samples) for organizing and preserving images, recordings, and data associated with genetic samples. This protocol can be used by individual researchers and institutions to organize their own collections, and it also facilitates submission of records to international data repositories such as VertNet. By contributing OMBIRDS to the research community as a free database tool that can be downloaded and adapted by researchers and institutions, we hope to encourage the collection of media along with genetic samples and to facilitate the archiving of these materials for their use in future research.

Keywords: blood samples, data archiving, DNA preservation, genetic samples, media specimen, museums, sound recordings, voucher specimen

Un appel à la conservation des images, des enregistrements et d'autres données en association avec des échantillons génétiques aviaires et introduction d'une solution : OMBIRDS

RÉSUMÉ

Beaucoup de recherches actuelles et historiques en ornithologie emploient des méthodes de capture-remise en liberté, ce qui procure diverses données et du matériel sur des oiseaux pour lesquels aucun spécimen entier n'a été récolté. Un spécimen génétique (ex.: sang ou plumes) est souvent récolté en plus des “spécimens virtuels”, comme des images et des enregistrements sonores. Cela constitue une importante source de matériel de recherche ainsi qu'une opportunité d'utiliser à chaque type de spécimens comme une source de validation des autres types. Malgré l'abondance de ces jeux de données et leur utilisation potentielle dans de futures recherches, la conservation de ces données et du matériel associé est une tâche à laquelle chaque chercheur est confronté, ce qui occasionne la perte de ce matériel de recherche au fil du temps. Pour promouvoir l'utilité à long terme des informations recueillies sur des milliers d'oiseaux capturés et remis en liberté à chaque année, nous présentons un protocole et un modèle de base de données (OMBIRDS: le musée en ligne des images, enregistrements et échantillons d'ADN d'oiseaux) pour organiser et conserver les images, les enregistrements et les données associées aux échantillons génétiques. Ce protocole peut être utilisé par des chercheurs individuels et des institutions pour organiser leurs propres collections et faciliter le transfert de leurs données à des banques de données internationales telles que VertNet. En offrant OMBIRDS à la communauté scientifique comme un outil gratuit de base de données qui peut être téléchargé et adapté par les chercheurs et les institutions, nous espérons encourager la collecte de matériel auxiliaire avec des échantillons génétiques et faciliter l'archivage de ce matériel pour leur utilisation dans de futures recherches.

Mots-clés: échantillons sanguins, archivage de données, conservation de l'ADN, échantillons génétiques, spécimen sur support visuel, musées, enregistrements sonores, spécimen de référence

Biodiversity loss is presently occurring at a rate unprecedented in human history, yet the technology available for the collection and analysis of biodiversity information is rapidly increasing. Digital photography, portable digital sound recorders, and advances in DNA sequencing have made it possible to routinely collect and study a wide variety of behavioral, morphological, isotopic, and genetic data from wild birds. As a research community, we have an opportunity to ensure the continuing availability of these data to future researchers, who can use them to verify our own scientific findings and to facilitate future studies—for example, on how biodiversity has changed over time. Tragically, however, vast quantities of scientific materials and data are currently being lost, as researchers retire, freezers stop working, and hard drives fail. An analysis by Vines et al. (2014) indicates that the odds of a dataset being accessible decreases by more than 17% per year. This estimate applies to data; the rate of loss of materials is likely much greater. In this era of climate change, rapid loss of biodiversity, and rapidly changing societal values, we feel that scientists, institutions, and governments should prioritize the safekeeping of research materials and data, so that they are accessible to future researchers.

The basic motivation for the long-term storage of collections of biological specimens in museums is to facilitate future research as well as the independent verification of scientific findings. Museums generally do an excellent job of preserving collected whole-body specimens, such as skins and skeletons. However, most ornithological research is now based on catch-and-release methods rather than whole-body collection (Figure 1). Such studies often involve the collection of genetic material (e.g., blood or feathers) along with occurrence data, morphological measurements, photographs, and/or sound and video recordings. Yet there is presently no well-established or broadly accepted protocol for archiving genetic material along with associated data. The majority of museums do not accept genetic samples without a specimen voucher (skin or skeletons, or some combination thereof). Existing databases dedicated to the collection of biodiversity sound, image, or video media (e.g., Macaulay Library, <http://www.macaulaylibrary.org>; Xeno-canto, <http://www.xeno-canto.org>; and Internet Bird Collection, <http://www.abc.lynxeds.com>) are not set up to simultaneously store and point to the existence of associated genetic material. Here, we call for the establishment of protocols for the long-term archiving of these diverse forms of materials and data in an integrated way, and we introduce and make available one particular protocol that we have implemented at the Beaty Biodiversity Museum at the University of British Columbia.

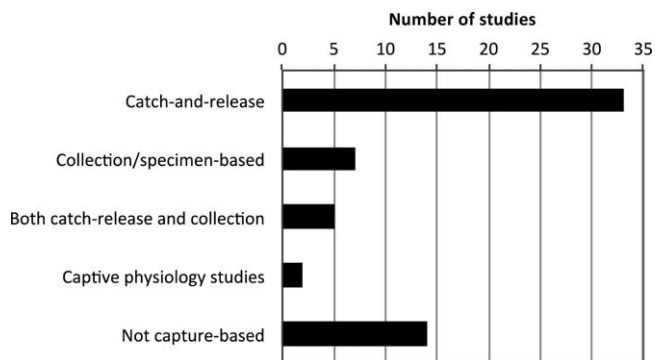


FIGURE 1. A survey of papers published in *The Auk* during the past year (late 2012–late 2013: issues 129[4] to 130[3]) reveals that the majority of studies (33 of 61) were based on catch-and-release approaches, and an additional 5 studies used catch-and-release and collection-based methods. For the above analysis, we excluded an additional 9 catch-and-release-based papers that were part of a special section on geolocators. Altogether, a total of 62–67% of studies employed catch-and-release approaches.

Value of Integrated Genetic and Media Specimens

Most scientific research relies on examining associations among multiple variables. By increasing the number of variables measured from each individual in a study, we increase the scientific value of each individual. For birds, we can collect genetic, behavioral, morphological, and other forms of information, and collecting a wide variety of information enables us to ask a greater variety of questions. For instance, are genetic and song variation associated in a putative contact zone between incipient species? Do genetically divergent forms differ in their plumage patterns? Do more variable songs correlate with larger body size? Do song and genetic diversity decline over time in a small isolated population? What parts of the genome correlate with song variation in the center of a hybrid zone between two species? Of course, collecting multiple types of information can be time-consuming, and researchers regularly make judgments about which information is feasible to collect and is most applicable to the questions being asked in a specific study. Currently, many researchers collect a variety of information and materials from each bird, but much of the complexity of these valuable resources will eventually be lost.

Although collection of whole-body specimens is essential for some research questions, many research projects use catch-and-release methods because of logistical, ethical, or permitting constraints, or because catch-and-release better enables the answering of questions that depend on tracking individuals over time (e.g., behavioral or ecological interactions, migratory routes, and return rates). Investigators conducting this style of research face two long-term challenges: (1) how to best organize their genetic material and associated data for their own use and

(2) how to ensure that these materials are accessible to future researchers.

With technological advances increasing the feasibility of making and using digital images and sound recordings in research, the bar has been raised on how much and what types of data can be gathered from individual birds. Even with traditional specimens (skin, skeleton, etc.), it is becoming common to capture photos, sound recordings, and/or other media prior to the collection of whole specimens (Bostwick and Scholes 2012, Webster 2013). The proper collection and storage of media is even more important for feather-only or blood-only samples, when whole specimens are not collected. In those cases, photographs and sound recordings can be used to help verify the sources of genetic samples (and vice versa). Many museum ornithologists may be wary of using photos and sound recordings as a sort of “voucher specimen” for a genetic sample, but surely all ornithologists can agree that using media to help verify sources of genetic samples is better than having nothing other than the notes of the collector (much current genetic research in birds is based on such a method). Furthermore, for many types of research (e.g., behavioral), media specimens associated with genetic samples are even more valuable than whole specimens.

It is important that these diverse types of research materials collected from individual birds be organized and safely stored as soon as feasible after being collected, since this will better ensure accuracy as well as accessibility. The optimum is a single data-entry process that begins in the field; continues during the lab analysis phase; functions as a research tool prior to publication; serves to store and preserve data; and streamlines the transmission of digital information to a data provider, ultimately notifying the research community of the existence of this material. Individual researchers can benefit from the increased organization and accuracy provided by such a protocol, as well as from the increased potential for citations and collaboration generated by making their materials available to the broader research community. Likewise, the broader community will benefit if all available media are linked to the corresponding genetic material and if the existence of these materials is advertised via international, multi-museum search engines (amalgamated data providers such as VertNet, Arctos, etc.).

OMBIRDS

We propose a new concept and protocol called OMBIRDS (Online Museum of Bird Images, Recordings, and DNA Samples) to promote and facilitate the preservation of linked genetic and media specimens. OMBIRDS simplifies the task of linking physical samples, media, and related data using an integrated database and standardized storage

protocols for data and media. Our goals in developing OMBIRDS are to (1) encourage the research and museum communities to collect, organize, and preserve linked genetic and media specimens; (2) provide a tool that researchers can use to organize their own specimens for their own use; (3) facilitate the establishment of institutional repositories of linked genetic and media specimens; and (4) provide an integrated protocol by which institutions can submit their OMBIRDS data to amalgamated data repositories such as VertNet (<http://vertnet.org>), GBIF (Global Biodiversity Information Facility, <http://www.gbif.org>), and Arctos (<http://arctos.database.museum>). Our vision is that OMBIRDS can be used throughout the entire data-collection and preservation process, starting at the level of the individual researcher, moving to the museum or institutional level, and ending at the level of international data repositories. These large data amalgamators collect and display information submitted by contributing institutions; OMBIRDS fills a different role, enabling researchers and institutions to organize information and materials in their own private database, submitting their data to larger databases if and when they choose to. We are contributing OMBIRDS to the research community as a database tool that can be downloaded and adapted by numerous researchers and institutions, rather than as a single database housed solely at our own institution. Of course, some institutions may choose to develop their own databasing protocols for integrating genetic and media specimens, and we applaud such efforts.

OMBIRDS, a user-friendly, one-time data-entry protocol and method of data organization and storage, is designed to aid and encourage researchers to store and preserve all information from an individual bird as a single database record. This includes the ability to store media within the database or as a referenced file stored on a server (thereby creating stable sites for media storage and access). Behind the user-friendly templates, all OMBIRDS data fields are configured using Simple Darwin Core terms to aid data transmission to other institutions. Simple Darwin Core is a standardized set of terms and definitions for sharing biodiversity data, from locality to morphometric information (Darwin Core Task Group 2009, Wicczorek et al. 2012). OMBIRDS Darwin Core–defined fields are stored within a relational database structure that relates tables of occurrences, events, locations, identifications, taxonomic information, and digital media and media metadata. The use of Darwin Core fields in OMBIRDS records will shorten the turnaround time between a museum accepting and adding media-vouchered specimens to their online database, because the majority of the labor-intensive data entry is already done. This, in turn, will shorten the time before the existence of these media-rich specimens is published via amalgamated database providers.

FIGURE 2. An example OMBIRDS record, showing a variety of data about an individual bird, photos of the bird, a Google Maps box with the occurrence location, and a sound recording (in this example, linked to a recording stored at the Macaulay Library, playable directly from the OMBIRDS record). More photos, sounds, and videos can be linked using the “Media” tab. Researchers are encouraged to create additional customized templates by adding data and/or media fields to tailor the data-entry template to their specific projects. The first syllable of the acronym, “OM,” is reminiscent of the first syllable of the Sanskrit mantra *Om Mani Padme Hum*; hence, the logo (at the upper left) is a bird containing the symbol for Om, meaning “wisdom.”

An example OMBIRDS database (Figure 2) and a free downloadable OMBIRDS data template are available at the Beaty Biodiversity Museum website (<http://beatymuseum.ubc.ca/ombirds>). The template runs on the widely used FileMaker Pro database program. Users must purchase their own licensed copy of FileMaker Pro, as they will be establishing their own OMBIRDS database. We welcome feedback from readers regarding the utility of the OMBIRDS template as well as suggested modifications. We encourage researchers and institutions to customize database layouts to create digital datasheets specific to their project needs. To do this, researchers need to duplicate the layout(s) downloaded from the OMBIRDS web page and then add or delete database fields as necessary.

We envision OMBIRDS being used by both individual researchers and museums and that its common use will facilitate the submission of data and research materials from individuals to museums, when and if individuals choose to do so. Researchers and institutions are responsible for the transfer, accessioning, and dissemination of data and genetic and media specimens according to institutional policies and practices, and for the establishment of relationships with appropriate data providers such as VertNet. Copyright issues should be discussed prior to accessioning of materials and data into official museum collections and data repositories; in most cases, researchers who submit materials to collections would be giving up control over those materials and agreeing that the sharing

of materials will be governed by the policies of the institution managing that collection. Many researchers will want to do this only after they have finished intensively researching the materials, often after they have published their findings. We anticipate, however, that given the substantial benefit of making the materials available (e.g., increased citation rates, increased potential for collaboration, and the satisfaction of having their materials available for future research), many researchers will contribute to such collections and databases. Additionally, funding agencies and journals increasingly encourage or require researchers to make data easily accessible. Readers who are curious about the evolving consensus regarding data use and data sharing may want to consult the policies of GBIF, a major biodiversity data provider that amalgamates information provided by various data publishers (see <http://www.gbif.org/disclaimer/datasharing>). We note that the standard practices of properly acknowledging the sources of research materials should apply to media and genetic specimens. In cases in which many specimens contributed by a particular collector are used in a research project, we suggest that it would be appropriate to contact that collector to discuss the possibility of coauthorship.

The OMBIRDS database template allows flexibility in terms of where media files are stored. At institutions that have FileMaker Pro linked to a server, the database enables the creation of stable URLs for media stored on that institution's server. Researchers may also enter into agreements with organizations such as the Macaulay

Library (Cornell Lab of Ornithology, Ithaca, New York, USA) for specialized media storage and access. Following the creation of stable URLs for all the media types and the insertion of these URLs into corresponding OMBIRDS records, an institution's OMBIRDS database is ready to be transmitted to VertNet using the GBIF Integrated Publishing Toolkit (Wieczorek 2011; for more information, see <http://www.vertnet.org/join/join.html>). All available media can be made available via VertNet, or researchers and institutions can choose to block some data or media as they wish. The use of OMBIRDS protocols does not require publishing through VertNet or other providers, though the publishing of OMBIRDS records is strongly encouraged.

At a minimum, an OMBIRDS record requires taxonomic identification down to the species or subspecies level; the date on which the bird was captured and/or media were collected in the field; and locality information composed of a descriptive location and a set of decimal-based geographic coordinates. OMBIRDS can store any number of media files and associate them with a particular occurrence record. However, a researcher may want to select a set number of images and other media files to facilitate ease of storage and encourage quality documentation. During this process, researchers can create an associated “bonus file” that is also linked to each database record for all the other images, songs, or videos of the bird that are not showcased in the OMBIRDS record. In addition to bird images, we recommend taking a habitat image, which should include the mist net if this was the mode of capture. Future researchers can utilize the habitat image to get a general impression of the study site or to look at changes in habitats and land use.

Example Archiving of Genetic and Media Specimens

We used the OMBIRDS protocol to transfer, store, and publish data and media from 73 MacGillivray's/Mourning Warblers (*Geothlypis tolmiei/philadelphia*) from a recently discovered hybrid zone as well as from allopatric areas (Irwin et al. 2009, Kenyon et al. 2011). Each blood sample was accessioned and given a University of British Columbia, Beaty Biodiversity Museum Cowan Tetrapod Collection (UBCBBM, CTC), catalog number. Associated field data were imported by mapping an Excel spreadsheet to Simple Darwin Core terms. Access profiles using usernames and passwords were assigned to individual users to protect the confidentiality of research prior to publication. Stable URLs of one habitat photo and up to 5 pictures from each bird were created on the UBC Zoology Server. We submitted song recordings to the Macaulay Library of Natural Sounds. After a license agreement was signed, the Macaulay Library hosted the files and created stable URLs for the song recordings of each bird. The

stable URLs created at both universities were inserted into the appropriate OMBIRDS database fields. The OMBIRDS records, including these stable URLs, were accessioned as part of the general Avian Research Collection at UBCBBM and transmitted to VertNet as part of the annual update of the UBCBBM CTC holdings (May 2013), thus publishing the existence of these media-vouchered blood samples to the entire scientific community.

The VertNet integrated portal (<http://portal.vertnet.org>) hosts specimen occurrence records across vertebrate taxa from multiple institutions. In addition to a Simple Darwin Core-based data record, the portal displays geolocated specimens in a Google Maps frame and links to externally hosted media. OMBIRDS records are integrated into the portal and take advantage of its new functionality (to see our OMBIRDS records, enter “ombirds” into VertNet's search box).

Conclusions

Widespread use of OMBIRDS protocols or similar approaches could dramatically increase the quality and quantity of museum-curated media-vouchered genetic material available to future researchers. This is true for traditional specimens (skin, skeleton, tissue, or some combination thereof) and when only blood or feathers are collected. We also note that, although we have focused here on the problem of avian genetic samples, these databasing efforts could be expanded to records of other taxonomic groups and situations in which physical samples were not collected—for instance, for encounters when photographs, sound recordings, location, and other data were recorded. In addition to facilitating future research, sharing is also likely to benefit current researchers who provide data and materials, by increasing citations of their original research and/or materials (Piwowar et al. 2007) and opening up opportunities for collaboration. Practices and procedures promoting the long-term preservation of data are increasingly encouraged in the fields of ecology and evolution (Whitlock et al. 2010); the archiving of genetic samples and associated media used in published research would serve a purpose analogous to that of the highly successful Dryad database (<http://www.datadryad.org>) for storing information associated with evolutionary and ecological research. We hope that the capability to preserve and store linked visual, acoustic, and genetic data from temporarily captured birds encourages both data sharing and the collection of more complete datasets, increasing the ability to use one form of data (e.g., photos) as a validation for another (e.g., genetic samples).

We feel that the best way to facilitate the creation of media vouchers is a single data-entry protocol such as OMBIRDS that aids in the entire process of data collection, analysis, and archiving. To obtain linked genetic

and media data of the highest quality, it is best if media vouchering is initiated close to the time of capture. To prevent data loss, all locality, morphometric, and media data collected (images, recordings, spectrograms, genetics, etc.) need to be stored as a relational database record, to create the most complete media voucher possible. We welcome the use and adaptation of the OMBIRDS protocol by researchers at other institutions and hope that the concept encourages researchers to organize and preserve the thousands of avian tissue and blood samples currently in the freezers of individual research groups. In this era of rapid biodiversity loss and climate change, the rapid establishment of repositories for genetic and media specimens at museums and the submission of research materials into those repositories would be a great contribution to future generations.

ACKNOWLEDGMENTS

We thank C. Cicero and J. Wieczorek at VertNet and M. Medler, G. Budney, E. Scholes III, and M. Webster at the Macaulay Library of Natural Sounds for their input and assistance. K. Bostwick, K. Winker, and other members of the AOU Collections Committee made valuable suggestions, as did P. Arcese, J. Irwin, W. Maddison, S. Otto, M. Whitlock, and two anonymous reviewers. Special thanks to D. Tan for creating the OMBIRDS logo.

LITERATURE CITED

- Bostwick, K. S., and E. Scholes III (2012). Bird Specimens 3: Alternative preparation types and media in modern avian collections. NAOC-V Using and contributing to avian collections workshop. Convener, I. Szabo. Beaty Biodiversity Museum. http://www.youtube.com/watch?v=jrlW8wKtmFU&list=PL7_GnRI3mdtq2i9ZwBAPXPmNKielUQ9Zq&index=3
- Darwin Core Task Group (2009). Darwin Core. <http://rs.tdwg.org/dwc>
- Irwin, D. E., A. Brelsford, D. P. L. Toews, C. MacDonald, and M. Phinney (2009). Extensive hybridization in a contact zone between MacGillivray's *Oporornis tolmiei* and Mourning warblers *O. philadelphia* detected using molecular and morphological analyses. *Journal of Avian Biology* 40:539–552.
- Kenyon, H. L., D. P. L. Toews, and D. E. Irwin (2011). Can song discriminate between MacGillivray's and Mourning warblers in a narrow hybrid zone? *The Condor* 113:655–663.
- Piwowar, H. A., R. S. Day, and D. B. Fridsma (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2:e308.
- Vines, T. H., A. Y. K. Albert, R. L. Andrew, F. Débarre, D. G. Bock, M. T. Franklin, K. J. Gilbert, J.-S. Moore, S. Renaut, and D. J. Rennison (2014). The availability of research data declines rapidly with article age. *Current Biology* 24:94–97.
- Webster, M. (2013). Ornithological specimens in the 21st century. Talk presented on August 17, 2013, at the American Ornithologists' Union Annual Meeting, Chicago, IL, USA.
- Whitlock, M. C., M. A. McPeck, M. D. Rausher, L. Rieseberg, and A. J. Moore (2010). Data archiving. *American Naturalist* 175:145–146.
- Wieczorek, J. (2011). The GBIF Integrated Publishing Toolkit User Manual, version 2.0. Copenhagen: Global Biodiversity Information Facility. <http://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes>.
- Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais (2012). Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7:e29715.