Journal of Mammalogy, 100(2):273–275, 2019
DOI:10.1093/jmammal/gyz059

A brief history of computerizing mammal collections and the role played by the ASM

Suzanne B. McLaren and Robert M. Timm*

Section of Mammals, Edward O'Neil Research Center, Carnegie Museum of Natural History, Pittsburgh, PA 15206, USA (SBM)
Department of Ecology & Evolutionary Biology and Biodiversity Institute, University of Kansas, Lawrence, KS 66045, USA (RMT)

* *Correspondent: btimm@ku.edu*

Beginning in the mid-1960s, curators at the largest mammal research collection in the world, the U.S. National Museum of Natural History, Smithsonian Institution, began the efforts of computerizing the data associated with their museum specimens (Squires 1966). Around the same time, the American Museum of Natural History, New York, initiated an index to the mammal collection using a mechanical indexing card system. Prior to these efforts to automate data retrieval, curators and researchers needed to either go directly to the specimens for information or search secondary sources such as hand-written, ledger-type catalogs or the hand-written or typed card catalog. The original motivation for capturing collection data was for management purposes (Van Gelder and Anderson 1967); however, the research value of ready access to accurate specimen data quickly became apparent. The first computer software application written specifically to input and manage natural history specimen data was developed at the Smithsonian Institution and called SELGEM, an acronym for SELf GEnerating Master (Creighton and Crockett 1971). The original development period was from 1970 to 1975, although improvements continued for several years. SELGEM and other computer programs were designed to streamline the process of manual cataloging of new specimens and data collection, while increasing the accuracy of data assembled, reducing repetitive work, and allowing for more flexible uses and retrieval of the available data. Specimen data were captured from the original field tags via keystroking the information onto a paper tape typewriter (Fig. 1); that paper tape was then fed into a mainframe computer. The SELGEM suite of computer programs was subsequently provided to the University of Kansas Natural History Museum (Lawrence, Kansas) and the Florida State Museum (Gainesville, Florida) by the Smithsonian Institution, along with the equipment and training of personnel; data capture began in 1973 at the University of Kansas following the Smithsonian's protocols. In the early 1990s, SELGEM data were converted to the next software application (Inquire), a text database, on an IBM mainframe computer, and retired in the early 2000s (D. Bridge, U.S. National Museum of Natural History, Smithsonian Institution, pers. comm., November 2018). The development of SELGEM protocols for data entry, and its rapid spread to other institutions provided a solid base for the rapid advancement of collection computerization.

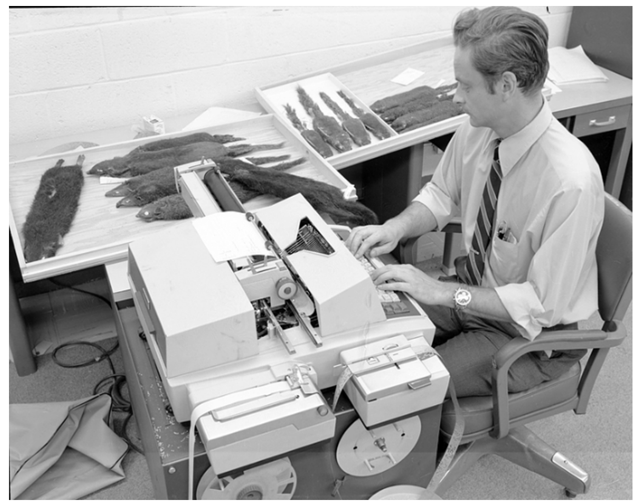

Figure 1.—Thomas J. McIntyre, National Museum of Natural History (NMNH), is operating a Frieden Flexowriter (model 2200 series), a paper tape typewriter. He is capturing data from specimen tags onto paper tape (seen to his left) for input into the Smithsonian's SELGEM computer system. McIntyre worked in the National Museum of Natural History, Automated Data Processing Program, for the Mammal Division from 1971–1974. Image taken November 1971. Courtesy of the Smithsonian Institution Archives Accession 11-009 (ID 71-510.tif) and T. J. McIntyre.

At the request of the National Science Foundation (NSF) in 1972, the American Society of Mammalogists established the Committee on Information Retrieval (now called Informatics Committee) and the Systematic Collections Committee, to assess the new and rapid expansion of collection databases and to seek guidance from a pool of well-informed members.

The Committee on Information Retrieval was established with a charge to examine the feasibility of developing a national data-retrieval system for Recent mammal collections and to develop funding for such a system; Sydney Anderson served as the first chair (Gill et al. 1987). At an ASM-organized workshop termed NIRM (Network for Information Retrieval in Mammalogy), guidelines, data standards, and suggested uniform data fields were established to support effective communication among institutions (Williams et al. 1979; McLaren et al. 1996). Professional societies dealing with other taxonomic groups later established guidelines following the ASM's lead. The early programs had limitations and operated on expensive mainframes, reducing their usefulness for smaller museums. These were replaced by more affordable minicomputers by the early 1980s. At the same time, commercial software such as dBASE for the DOS platform, FileMaker for the Macintosh platform, and later Microsoft's Access and Excel, and others came into use. This provided greater flexibility and data accessibility to collection managers and researchers, allowing them to capitalize on expertise available at the home institutions. As a result, various museums worked to develop their own software (American Museum of Natural History; Carnegie Museum of Natural History, Pittsburgh; Field Museum of Natural History, Chicago; Royal Ontario Museum, Toronto, among others; i.e., see McLaren et al. 1985; Arrigo and Timm 1987). However, with the various software systems being used, there was little sharing of data and expertise among institutions; each institution was essentially a stand-alone system.

To facilitate networking among research collections and provide support for georeferencing collection data, curators at two of the largest university mammal collections were asked to submit a proposal to the NSF's Biotic Resources Program. Philip Myers at the University of Michigan's Museum of Zoology, and James L. Patton at the University of California, Berkeley's Museum of Vertebrate Zoology, were funded in 2001 to develop the protocols and provide support for computerizing and georeferencing specimen localities in their extensive mammal collections. Following the success of the initial project, a collaboration among 17 mammal research collections in the U.S. and Canada was funded by NSF to expand and improve upon their databases with a shared effort at georeferencing. Staff from these collections were brought together to discuss how a distributed database would work during special sessions at the Annual Meeting of the ASM prior to NSF funding. The project called MaNIS (an acronym for Mammal Networked Information System) provided for the development of an integrated network for distributed databases of mammal specimen data. Project objectives included facilitating open access to combined specimen data from a web browser, enhancing the value of specimen collections, making the best use of curatorial resources, providing a design paradigm that could be adopted by other disciplines with similar needs, and opening the door to researchers seeking specimen record data. Additional collections and other vertebrate groups and their home institutions were subsequently funded by NSF. As of October 2013, the MaNIS community and data sources were subsumed in VertNet (http://vertnet.org), but the still-viable website provides details on best practices in georeferencing, a georeferencing calculator, a time-line of MaNIS events, and the names and acronyms of the 17 original collections involved in the project (http://manisnet.org/index.html; accessed 6 February 2019). Stein and Wieczorek (2004) provided a history of the MaNIS project. Software to search museum specimen records now includes GBIF, iDigBio, and VertNet. Researchers have found such integrated databases to be an extremely valuable data source facilitating and enhancing research, education, conservation, and public health.

Due to the rush to enter specimen data into electronic databases, the time-consuming work of data verification lagged behind the data capture process, leading to concerns by collection staff that researchers were not critically evaluating data prior to use (McLaren and Timm 1988). The problem of lack of rigor and how to evaluate the strengths and weaknesses of computer-based specimen data continues to be actively debated today (see Gutiérrez 2016; Bloom et al. 2018; and references therein).

Today researchers can access specimen data from most mammal collections throughout the world from their offices via their web browser, and this is true also for birds, reptiles and amphibians, many botanical collections, and other taxa. The first attempts with 1970s and 1980s computers and databases seem especially primitive by today's standards, but the technology and standards rapidly improved. In addition to serving as a valuable resource to the home institutions in managing their collections, unforeseen applications for specimen data continue to be realized. The rapidly expanding field of modelling distributions, and especially how climate change impacts the earth's flora and fauna, is one of the many research areas made possible using specimen-based data. Today the next generation of data management software systems (for example, ARCTOS, EMu, Specify, and Access) are used across many collection disciplines worldwide, and internet connectivity provides a mechanism for rapid support among users. More than fifty years ago, computerization of the world's biotic collections was spearheaded by mammalogists and the American Society of Mammalogists, and those efforts provided the foundation that helped make exciting new research and conservation efforts possible.

Acknowledgments

We thank D. Bridge and T. J. McIntyre for freely sharing their knowledge of ASM and USNM history with us, significantly contributing to the details presented here. J. H. Beach provided valuable comments on an earlier version of this manuscript. M. G. Girard's masterful efforts in improving the figure used herein are most appreciated. We acknowledge Donald F. Hoffmeister, the ASM's first and long-time Historian (1966–1995) and Past President (1964–1966), for his efforts in documenting the fascinating history of our Society. Following his example, in our roles as the ASM's current Historian (Timm) and Assistant Historian (McLaren), throughout 2019 we will attempt to provide interesting

aspects of the Society's history in the *Journal* to complement our 100th anniversary celebration.

Literature Cited

Arrigo, R. D., and R. M. Timm. 1987. Minicomputer system at Field Museum of Natural History. Pp. 121–128 in Mammal collection management (H. H. Genoways, C. Jones, and O. Rossolimo, eds.). Texas Tech University Press, Lubbock.

Bloom, T. D. S., A. Flower, and E. G. DeChaine. 2018. Why georeferencing matters: Introducing a practical protocol to prepare species occurrence records for spatial analysis. Ecology and Evolution 8:765–777.

Creighton, R. A., and J. J. Crockett. 1971. SELGEM: A system for collection Management. Smithsonian Institution Information Systems Innovations 2(3):(2 pages), 1–24, Appendix A: pages 1–3.

Gill, A. E., D. P. Christian, J. N. Layne, J. S. Millar, and J. L. Patton. 1987. Roles of standing committees of the American Society of Mammalogists. Ad hoc Committee to Evaluate Standing Committees. Journal of Mammalogy Supplement 68(1):1–13 + table.

Gutiérrez, E. E. 2016. Ecological niche modelling requires real presence data and appropriate study regions: a comment on Medone et al. (2015). Philosophical Transactions of the Royal Society B. 371:20160027. DOI: 10.1098/rstb.2016.0027

McLaren, S. B. (Chair). 1985. Survey report on computerized information retrieval in mammal collections. Committee on Information Retrieval, American Society of Mammalogists.

McLaren, S. B., P. V. August, L. N. Carraway, P. S. Cato, W. L. Gannon, M. A. Lawrence, N. A. Slade, P. D. Sudman, R. W. Thorington, Jr., S. L.Williams, and S. M. Woodward. 1996. Documentation standards for automatic data processing in mammalogy, Version 2. American Society of Mammalogists, Lawrence, Kansas.

McLaren, S. B., and R. M. Timm. 1988. Guidelines for computer usage of computer-based collection data. Journal of Mammalogy 69:217–218.

Squires, D. F. 1966. Data processing and museum collections: a problem for the present. Curator 9:216–227.

Stein, B. R., and J. Wieczorek. 2004. Mammals of the world: MaNIS as an example of data integration in a distributed network environment. Biodiversity Informatics 1:14–22.

Van Gelder, R. G., and S. Anderson. 1967. An information retrieval system for collections of mammals. Curator 10:32–42.

Williams, S. L., M. J. Smolen, and A. A. Brigida. 1979. Documentation standards for automatic data processing in mammalogy. The Museum of Texas Tech University, Lubbock, Texas.