

MAMMALS OF THE WORLD: MaNIS AS AN EXAMPLE OF DATA INTEGRATION IN A DISTRIBUTED NETWORK ENVIRONMENT

BARBARA R. STEIN AND JOHN WIECZOREK

Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, CA 94720

Abstract.—Natural history collections are the authoritative source of knowledge about the identity, evolutionary relationships, and attributes of species with which we share this planet. As such, collections of research specimens play a central and critical role in the conservation and sustainable use of biodiversity. The potential contribution of specimen data to systematic, genomic, and ecological analyses is enormous, and will be orders of magnitude greater when information is made easily accessible via distributed networks compared with stand-alone database systems in use up to the present. The Mammal Networked Information System (MaNIS) is a distributed database network that permits participating institutions to provide web-based global access to their collections data for research, education and informed decision-making. The simplicity of the network's design ensures that any institution wishing to join MaNIS may do so at relatively little cost and with relatively little technical expertise. Although development of MaNIS and its underlying architecture relied on a number of key programming tasks and innovations, much of what the project can offer at this pivotal juncture is insight into its approach and a template by which other disciplines can engage in a similar process with equal success.

Key words.—mammals, natural history museums, biodiversity data, distributed data networks, MaNIS, DiGIR

The desire to link and retrieve electronic data from geographically distributed natural history collections to increase their effective use in research, conservation, and education is not new, but successful attempts to achieve that objective have only emerged within the last decade. The first distributed query system (1993) for natural history collections was the National Science Foundation (NSF)-funded FishGopher project¹. In contrast, a centralized data warehouse approach was taken by the Neodat II project in 1997². This was followed by implementation of the Z39.50 profile for distributed natural history collections data (ZBIG), which debuted in 1998 with the Distributed Database of North American Bird Data (Peterson *et al.*, 2003). At approximately the same time, a consortium of Mexican and foreign institutions made their combined specimen information available online through REMIB, the Mexican Network of Biodiversity Information³ using TCP/IP sockets to ensure safe and efficient data transmission. An effort to exploit the Z39.50 technology in another discipline followed in 2000 with a joint collaboration between the ichthyology and marine biology communities in development of FishNET (Vieglais *et al.*, 2000; Wiley *et al.*, in prep.). Use of the Z39.50

protocol and TCP/IP sockets represented an advance over the centralized data warehouse concept by using a standard for distributed information retrieval that simultaneously accessed data directly from institutional databases.

A serious commitment to develop an interdisciplinary standard for unified interoperability among natural history databases, as opposed to continued insular parallel developments within taxonomic disciplines, was first voiced at the annual meeting of the Taxonomic Database Working Group (TDWG) in Frankfurt, Germany in November, 2000⁴. A committee was formed to create guidelines that would support such a collaboration and it was agreed that interoperability would benefit from the adoption of a network architecture that not only used a mainstream transport protocol (HTTP) and document structure (XML), but that also allowed user communities to define the structure of the data to be shared without affecting the protocol or existing software. What emerged from subsequent deliberations is a highly successful, cooperative, open source, international development effort. The resulting protocol, Distributed Generic Information Retrieval (DiGIR)⁵, is designed to support unified queries to geographically-distributed providers of data via one or more portals—software installed in

¹ <http://www.as.ua.edu/biology/uaic/fishgopher.html>.

² <http://www.neodat.org/>.

³ http://www.conabio.gob.mx/remib_ingles/doctos/remib_ing.html.

⁴ <http://www.tdwg.org/news2001.html>.

⁵ <http://digir.net>.

conjunction with a web server that manages the connections to, constructs and sends requests for data to, and receives and processes responses from DiGIR providers.

The Mammal Networked Information System (MaNIS), an NSF-funded project⁶ consisting of a consortium of seventeen North American mammal collections, was developed simultaneously with DiGIR and was the first functional implementation of a DiGIR-based network. The readiness of the mammal community to share and benefit from the power of the combined data in their institutional collections became a major force driving development of DiGIR. Functional implementations of both provider and portal software were needed by the DiGIR project to demonstrate the utility of the protocol. A set of natural history concepts was also needed so that collections could map information in their databases to a well-defined semantic standard. These same requirements were central to MaNIS for the creation of a scalable network.

From an informatics perspective, the mammal community was well suited to become involved in development of a distributed database network. The discipline encompasses a relatively stable taxonomy; the taxa are well-known at the species level and they follow a generally-accepted taxonomic authority that has been rendered into electronic form (Wilson and Reeder, 1993). This volume, *Mammal Species of the World*, is used as the basis of the taxonomic entries in the Integrated Taxonomic Information System (ITIS⁷) and Species 2000⁸. In addition, mammal collections in North America have achieved a high degree of computerization of their specimen data (Hafner *et al.*, 1997). Under the auspices of the Committee on Information Retrieval of the American Society of Mammalogists (ASM), documentation standards for data processing in mammalogy (McLaren, 1999) were developed more than two decades ago (Williams *et al.*, 1979) and by late in 2000 had been implemented by essentially all North American institutions that house mammal collections, regardless of size. Early guidelines for the use of computer-based collection data (McLaren, 1988) also highlight the preparedness of this taxonomic discipline to meet the social and technical

challenges of establishing a distributed database network.

Development of DiGIR, and in turn the MaNIS network⁹, relied on a number of key programming tasks and innovations, including designing conceptual schemas (for specimen data as well as taxonomic and geographic indexes) to assist in data discovery, writing migration scripts to periodically update data in MaNIS repositories, designing worldwide web interfaces and data portals, creating tools for providers to monitor usage, responding to diagnostic messages from providers, allowing for data caching, dealing with unresponsive providers, and creating a provider installation package for institutions that were not funded through the original NSF award. Independent development of web-based tools and guidelines to facilitate efficient georeferencing of specimen localities¹⁰ was of equal import to make these data valuable to a much broader segment of the research community (ecologists, biogeographers, conservationists, educators, etc.). Although a great deal of technical development was required to create the MaNIS network, it can be strongly argued that the sociological issues involved in successfully developing a distributed database network within a community of scientists and institutions eclipsed any of the technological problems that had to be overcome. Data and software aside, much of what MaNIS can offer at this pivotal juncture is insight into its approach to the project and a template by which other disciplines can engage in a similar process with equal success.

MANIS OBJECTIVES

Genesis of the MaNIS network can be traced directly to a symposium held at the ASM annual meeting in Seattle, Washington in 1999. Entitled, “Emerging Database Technologies”, the symposium included a presentation by A. Townsend Peterson (University of Kansas Natural History Museum and Biodiversity Research Center) about the Z39.50-based avian network and the vast array of research and conservation questions and applications to which the power of such combined specimen data could be applied. By the close of the Seattle meeting, representatives from 17 North American mammal collections agreed to collaborate on the development

⁶ <https://www.fastlane.nsf.gov/servlet/showaward?award=0108161>.

⁷ <http://www.itis.usda.gov/>.

⁸ <http://www.sp2000.org/>.

⁹ <http://elib.cs.berkeley.edu/manis>.

¹⁰ <http://elib.cs.berkeley.edu/manis/search.shtml>

of such a distributed database network. In order to keep the project affordable and manageable in a three-year funding period, no further recruiting was attempted.

The initial clarity of the objectives of MaNIS and the concrete benefits articulated to and recognized by the participants were directly responsible for development of a successful proposal to NSF. It was noted from the outset that design of the network needed to benefit the participants as well as the larger user community. Recent closures and significant reductions in programs and operating budgets at many university and free-standing museums was, and still is, a reality that needed to be addressed when asking curators and their staff to make a significant commitment of time and effort to the MaNIS project. When crafting the project proposal, it was acknowledged that institutions would be unwilling and/or unable to relinquish their current in-house database management systems in order to participate in MaNIS, that they would have to be able to document network use of their collections, and that collections support and new hardware would enhance their ability to participate in the project, as well as their standing with institutional administrators. Equally important, the design of the network would have to ensure that institutions retain control over which data were accessible to address concerns about the security of rare species and the intellectual property rights of institutional data providers (Brooke, 2000; Graves, 2000).

Developing an approach to the project that addressed these issues was essential. Moreover, it was generally felt that MaNIS would afford each of the participating institutions and their collections heightened visibility within the larger research community. For many of the original participants, MaNIS presented the first opportunity to make their specimen data directly accessible via the Internet. Increased visibility and ease of data access via the network would, in turn, validate the importance of the collections, with greater institutional support viewed as a potential outcome. From the participants' viewpoint, the project simply reinforced the unambiguous value they placed on their own collections for research, conservation, and education.

IMPLEMENTATION ISSUES

In addition to the clarity of objectives noted above, it was recognized in the earliest planning stages of MaNIS that the network architecture had to

be simple, low cost, and require minimal maintenance. Other constraints driving project design included no visible long-term support for the network or its participants, known opposition within the community to centralization of operations, and uncertain availability of in-house technical expertise to maintain institutional systems after the funding period ended. In addition, given an articulated commitment to develop an interdisciplinary standard for unified interoperability among natural history databases, every attempt was made to conceive a design that could be easily adopted by others with similar needs.

The resulting MaNIS network architecture is novel in two respects. First, each MaNIS data provider automatically maintains summary data (counts of specimen records indexed by taxonomy and geography), in addition to specimen data from its institutional database. Second, the configuration of the publicly accessible repositories is optimized for query performance rather than for data management, for which the institutions' in-house databases are optimized. Data are filtered and standardized at each institution in automated periodic migrations from curatorial databases to the MaNIS data provider repositories. This design allows institutions to retain control over public access to their data without the need to create new structures, or change original data for public consumption, in their curatorial databases. Use of replicated databases also protects curatorial databases from increased traffic and unsolicited intrusion. Given the age of many curatorial databases, this design element was considered essential. In addition, the network will continue to function even if a curatorial database experiences problems or requires maintenance.

To keep the broader natural history community apprised of our progress, a project web site was established and all relevant documents were posted¹¹. The site has made a much more interdisciplinary group of researchers aware of the project than was initially envisioned. Thus, an educated community of users and enthusiasts has been slowly gaining ground through workshops, training sessions and presentations at scientific conferences. A page on the MaNIS web site summarizes the most notable of these events¹², providing a limited chronology of the evolution of distributed data networks within the

¹¹ <http://elib.cs.berkeley.edu/manis/Documents.html>.

¹² <http://elib.cs.berkeley.edu/manis/Events.html>.

natural history community, as well as the MaNIS project itself.

STANDARDS ARE PARAMOUNT

The importance of standards in all phases of development of MaNIS cannot be overstated. Because of the informatics groundwork that had been laid down by the mammal community prior to the initiation of MaNIS, both the semantics (the standards expressed in a conceptual schema or map of database concepts and their relationships¹³), and data standards, (the recommended standard vocabulary for data content (McLaren, 1999)) were adopted by the group with relative ease. Such agreements are essential in any federation to be sure that fields are populated with consistent values having the same meaning across collections. The quick adoption within the context of MaNIS does not trivialize the necessity for other communities to formulate comparable standards if they do not exist prior to embarking on a similar project. Furthermore, there are benefits to cross-disciplinary standards that should also be considered. For example, a set of common core concepts between mammalogy and parasitology would allow information from both disciplines to be accessed simultaneously and shared without confusion.

Within the mammal community, the existence of a relatively stable and accepted taxonomy and the prior existence of data content standards meant that the specimen data housed in participant collections were already largely consistent across institutions and in suitable electronic form at the time of proposal preparation. Of equal import, this meant that subsequent discussion of a data exchange standard (the federated conceptual schema) precipitated relatively little debate. Only slight modifications to the initial draft were required, because the concepts agreed upon simply mirrored those in the participants' institutional databases and had long-standing acceptance within the discipline as a whole. This standard is now recognized as an extension of Darwin Core Version 2 (DwC2)¹⁴, a profile describing the minimum set of standards for search and retrieval of natural history collections and observation databases.

The advantage of having standards established within the mammal community was also reflected in

the ease with which the participants were able to agree upon and adopt standards for georeferencing the specimen localities in their collections, i.e., assigning geographic coordinates and maximum error distances for those coordinates to locality descriptions. As stated above, much of the value inherent in specimen data is dependent upon the presence of accurately georeferenced locality information. This is particularly true if the data are to be used for modeling and predictive analyses (e.g., in the fields of conservation biology, ecosystem monitoring, and disease tracking). Less than 26% of the roughly 1.4 million mammal specimens housed in participating museum collections contained coordinate data in conjunction with specimen localities at the time of proposal submission. Hence, coordinating georeferencing activities was a major focus of activity for the project. At the outset of the MaNIS project, there were no established standards for georeferencing descriptive locality data. The development of georeferencing guidelines¹⁵ and tools was critical to the success of the collaborative approach that was adopted, and both have proven useful to other initiatives as well.

COLLABORATION

Georeferencing

Collaboration, in addition to standards, has also been key to the success of MaNIS. On one hand, it has resulted in cost efficiencies and economies of scale that would not otherwise have been realized. On the other hand, the project has recognized tremendous benefit from having a community of individuals available to address and propose solutions to problems as they arise. Both of these factors can be demonstrated most clearly with examination of the collaborative georeferencing activities that lie at the heart of the project.

Getting participants to recognize the value of collaborative georeferencing did not require much effort. Yet, on a specimen level, there was immediate consensus that georeferencing the locality from which every organism in a medium-sized or large collection had been collected was a daunting and seemingly overwhelming proposition. With every institution facing similar data problems (outdated place names, vague or ambiguous localities, specific localities mismatched with higher level geographic attributes) and limited resources, even with NSF

¹³ <http://elib.cs.berkeley.edu/manis/darwin2ConceptInfo030315jrw.htm>.

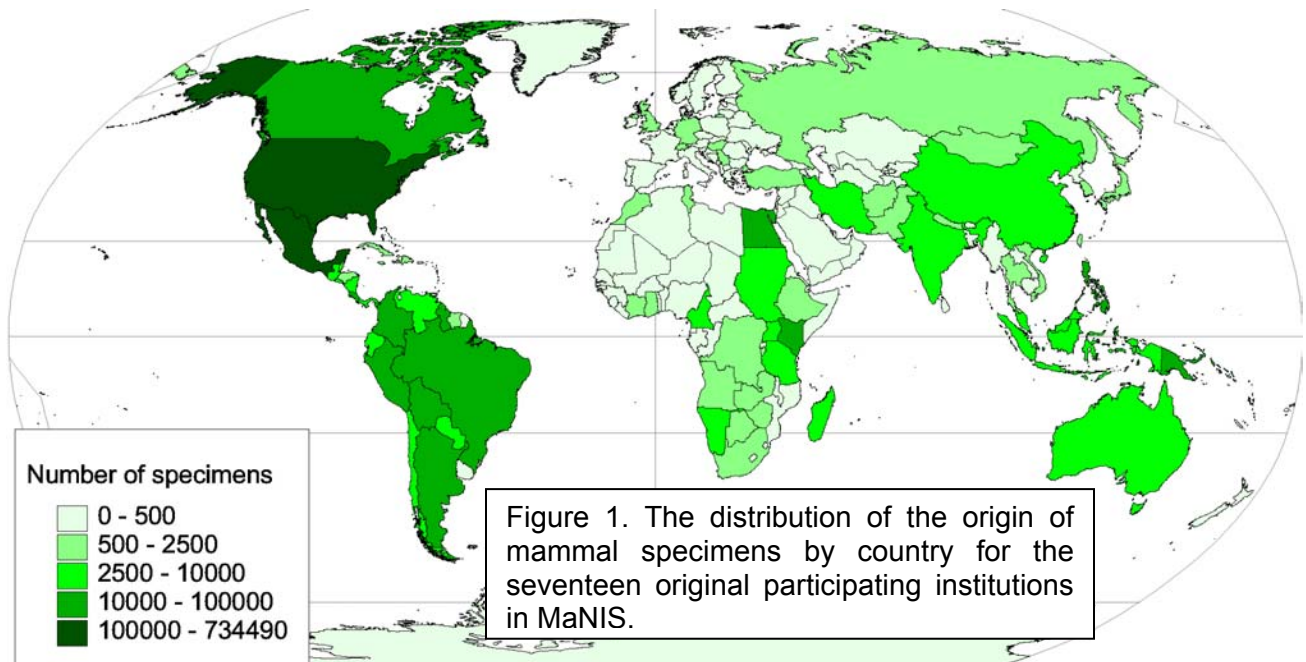
¹⁴ <http://tsadev.speciesanalyst.net/documentation/ow.asp?DarwinCoreV2>.

¹⁵ <http://elib.cs.berkeley.edu/manis/GeorefGuide.html>.

support, duplication of effort had to be precluded. Although the primary goal of each institution was to make sure that all of the localities from its own collections were georeferenced at the conclusion of the project, a plan to share the georeferencing workload among participants was readily adopted.

Perhaps the single most important innovation in the project was the creation of the MaNIS georeferencing gazetteer, which contained all unique specimen collecting localities (296,737) for the 1,367,627 specimens from the seventeen original institutions in the project (Figure 1), including the 64,073 localities for which geographic coordinates were already provided. This gazetteer was created by combining locality data from all participating institutions at the outset of the project into a single gazetteer database with an online query interface for browsing and downloading tab-delimited locality records¹⁶. The gazetteer was used as the basis for collaborative georeferencing, whereby each

institution reserved geographic areas in which they had particular interest, map resources, or expertise¹⁷. Having claimed a geographic region for georeferencing, an institution would query the gazetteer for all of the localities from that region, download them, and georeference all of the localities for all of the participating institutions. By having one institution georeference the localities of all institutions for a given region, tremendous economy of scale, uniformity, and use of expertise were achieved. It also meant that each of the participating institutions did not need to possess or acquire geographic resources (e.g., maps, gazetteers, atlases) for the entire world. This proved to be both a cost savings and a practical solution to a complex problem.



¹⁶ <http://elib.cs.berkeley.edu/manis/search.shtml>.

¹⁷ <http://elib.cs.berkeley.edu/manis/Checklist.html>.

When georeferencing of a geographic region was finished, the data set was uploaded, standardized for consistency of values (datums, units, etc.), checked for completeness, assessed for compliance with the georeferencing standards, and underwent geographic validation to ensure that the coordinates provided were in agreement with the large-scale geography describing the locality. The geographic validation process revealed that ca. 4% of georeferenced localities required further attention. Of those requiring further attention, roughly half were found to be correct, but had made it onto the list for additional examination due to inaccuracies in the boundary layer data (e.g., county shape files). The high rate of accuracy achieved using the guidelines and collaborative georeferencing are a testimony to the merit of the approach.

Looking forward, georeferencing in the MaNIS project has defined a baseline against which to compare the accuracy and speed of automated georeferencing techniques, and from which taxonomic specialists can make further refinements as time and money permit. Where digital maps were available for a geographic region, the mean (± 1 SD) georeferencing rate was 16.6 (± 8.3) localities per hour ($n = 14$ data sets from 14 institutions). The mean georeferencing rate for regions where printed maps were used instead of digital media was 9.6 (± 6.8) localities per hour ($n = 39$ data sets from four institutions). These rates include the determinations of both coordinates and uncertainties, with full documentation (Wieczorek, *et al.*, 2004). The metadata for uncertainties and documentation were found to take roughly 30%, on average, of the time required to georeference a given locality. Yet, without these additional data, it would not be readily possible to filter georeferences for suitability for a given line of research. Thus, it was deemed essential to provide these additional data. This view was later adopted by the TDWG Spatial Standards subgroup in its definition of an “acceptably georeferenced specimen record” (Frazier, 2002).

Collaborative georeferencing was initially an act of faith on the part of the participants, faith that time spent georeferencing localities from another institution would be rewarded by similar behavior on the part of one’s colleagues. Although unanticipated, this course of action was richly rewarded through the attraction of georeferencing partnerships and collaborations with organizations outside the 17 participating institutions. This produced immediate,

tangible benefits for the project, as well as the larger community. Shortly after MaNIS activities commenced, the lead programmer was contacted by Mexico's National Commission for the Understanding and Use of Biodiversity (CONABIO) with an offer to assist in georeferencing Mexican localities, using their own resources, for specimens housed in participating collections. Given that the 17 participating institutions housed 172,596 Mexican specimens from 30,059 distinct localities, this offer was accepted with alacrity. By adhering to the MaNIS georeferencing guidelines, complete georeferencing of those localities occurred in record time and with unparalleled expertise. In turn, CONABIO was able to greatly enhance its knowledge of historical distributions for Mexican mammals, which enables the organization to address ongoing species conservation and land management issues more accurately now and in the future.

Creating the Network Software

Georeferencing locality data in participant collections was only one of three major aspects of development of the MaNIS network. The other two aspects included creating the network software and connecting the participant institutional databases to the network. Though connecting institutions could not be initiated until a functional network was in place, the creation of the network software was another example of a collaborative effort that has benefited both MaNIS and the larger biodiversity informatics community.

As noted above, development of a reference implementation of the DiGIR software was, in part, driven and influenced by the requirements of the MaNIS participants. Yet, because of its international collaborative origins, the software’s diverse capabilities exceed what MaNIS programmers would have been able to achieve on their own within the limited period of awarded funding. By distributing responsibility for development of various software modules and functions (provider, portal, and network registration) among a diverse group of “biologists cum software developers” (see Acknowledgments), the project has been able to avail itself of distributed expertise in excess of funds available to any one interested party. Moreover, the collaborative nature of DiGIR development has resulted in widespread adoption of the protocol because its generic character is easily adapted to varying community standards

and disciplines and does not reflect the imprint of any one group.

Benefits of Collaboration

For collaboration to be successful among individuals at institutions with similar high-level goals but diverse circumstances and constraints, it is important to recognize what drives such parties to cooperate. From the perspective of the MaNIS participants, we would argue that a heightened sense of community and recognition of the enormous value of their combined data – “the whole” – has emerged. Simultaneously, both large and small collections are now recognized for their respective contributions to that whole. Appreciation of what can be gained through collaboration has replaced a sense of competition that is seemingly pervasive in scientific communities.

Education has also emerged as a valuable product of this collaboration. Understanding of technical issues has increased enormously among the participants. With respect to georeferencing, for example, the nature and value of the geodetic datum (the geometric description of a geodetic surface model such as NAD27, NAD83, and WGS84), errors, and extents has led to a desire among the participants to modify their current database management systems to accommodate these fundamental concepts. This will mean that new specimens added to their collections will have locality data of quality comparable to that which has been achieved in the georeferencing for the MaNIS project.

If one path to success is through collaboration, how has this been achieved in this project? Foremost was the establishment of trust among the participants with the programmer and with principles involved. MaNIS was extremely fortunate in having a lead programmer who possessed a keen understanding of and experience with mammalian biology, curatorial practices, specimen data, museum databases, and field work. Meetings between the project participants and the programmer were held at the annual ASM meetings throughout the grant period. Those meetings were designed to excite participants about each upcoming phase of the project, demonstrate new tools that would be made available to the project, answer questions face-to-face, and allow participants to interact and develop a rapport with the individual to whom they addressed their technical

questions or were forced to reveal the dark secrets of their collection data during the course of the project.

It was also made clear to participants from the moment of project conception that institutions would retain control over their data and over the entire project process. Accordingly, a listserv was set up to facilitate grant submission. Ultimately, this tool proved most valuable in creating a sense of community. By establishing a subscription list rather than a public list, all interested parties could take part in discussions, whether or not they were part of the original proposal. However, funded participants felt comfortable in posting questions about their data to the list, knowing that those who would read their postings probably shared their particular problem, had expertise in their problem area, or at least had a genuine interest in helping to solve problems. Because all MaNIS business was conducted openly on the list, and all decisions among participants were reached by consensus, each of the institutions was an equal player, regardless of size or perceived import of their collections. The enthusiasm among the participants was instrumental in making management of the project tractable and should not be underestimated when undertaking a project of this nature. Because of the cooperative fulfillment of shared goals, in the final year of the grant period, the project is on target for successful completion and there have been no unexpected or unwanted obstacles.

IMPACT

As stated in the NSF project proposal, the goals of the MaNIS project were to facilitate open access to combined specimen data from a web browser, enhance the value of specimen collections, conserve curatorial resources, and use a design paradigm that could be easily adopted by other disciplines, all while adopting mainstream transport protocols, avoiding the long-term, external maintenance of a network and centralized data management, and exercising fiscal economy. Ironically, remaining focused on those goals has presented the greatest challenge to the project due to the unexpected “ripple effect” that MaNIS has had. Given the enormous interest that the project has generated, it is increasingly difficult for participants not to be distracted by the burgeoning number of projects related to development of DiGIR and creation of the

MaNIS network (HerpNet¹⁸; OBIS¹⁹; SpeciesLink²⁰; GBIF²¹; ORNIS; BioGeomancer²²). From a programmer's perspective, the temptation to become involved in this flurry of activity is almost irresistible. From an institutional perspective, external pressures to divert limited resources or overextend collections staff in the name of ancillary projects that proffer immediate heightened visibility is an ongoing challenge.

A functional MaNIS network was first publicly demonstrated in June 2002 at the ASM annual meeting in Lake Charles, Louisiana. Later that month, the network was again demonstrated at the Global Biodiversity Informatics Facility (GBIF) Data Access and Database Interoperability Scientific and Technical Advisory Group in an invited presentation entitled, "The Mammal Networked Information System (MaNIS), powered by Distributed Generic Information Retrieval (DiGIR)." A third presentation of MaNIS was given in the Biodiversity and Biocomplexity Informatics Panel at the July 2002 Joint Conference on Digital Libraries in Portland, Oregon. At the 2003 annual meeting of the ASM in Lubbock, Texas, a symposium was held featuring five presentations covering topics relevant to the symposium title "The MaNIS Project: The Development and Applicability of a Mammal Collection Database Network." These demonstrations helped establish standards and allowed for community input throughout the development process.

There has been an enthusiastic reception to MaNIS well beyond the scope of its participating institutions. As a reference implementation of DiGIR, MaNIS has been a highly visible pilot project for a variety of international consortia, such as *The Species Analyst* (TSA²³ Peterson *et al.*, 1998), Australia's Virtual Herbarium²⁴, Landcare Research of New Zealand²⁵, the European Natural History Specimen Information Network²⁶, and Centro de Referência em Informação Ambiental (CRIA²⁷) in Brazil. Among North American distributed database

initiatives, HerpNet and the forthcoming effort on the part of the avian community, ORNIS, have both cited MaNIS as their model and have based their network architectures on DiGIR. In addition, there has been keen interest in the use of DiGIR as an elegant, inexpensive solution to the problem of web data access and database interoperability for local consortia of natural history data providers. Among these are the University of California (UC) Berkeley Natural History Museums and the Association of Biological Collections at UC Davis. In early February 2004, the Global Biodiversity Information Facility (GBIF²⁸) announced the public debut of its global DiGIR-based network, including the MaNIS participating institutions among the data providers of nearly 10 million specimen records.

The MaNIS project contributed both financially and academically not only to the development of the DiGIR protocol but also to the Darwin Core Version 2 conceptual schema for natural history collections data. The Darwin Core Version 2, an elaboration and reworking of the Darwin Core developed for *The Species Analyst* in 1999, captures the roughly 50 most prevalent biodiversity information concepts in common among natural history collections. This core provides a shared, standard information domain so that biodiversity information can be queried across multiple disciplines while ensuring compatible results.

Many of the questions posed during the development of the MaNIS network, and many requests for use of the data even before the network was functional, have shed light on new and potential research applications to which these data can be applied. Although museum curators have tirelessly and cogently articulated the value of their collections and associated data for decades, creation of the network has given substance to those words in a manner that is immediately accessible and understandable to administrators, colleagues, and the general public as never before. This has led to a renewed and enhanced appreciation of these invaluable resources and bodes well for their future support and the conservation of biodiversity on Earth.

ACKNOWLEDGMENTS

With support from the National Science Foundation, the following natural history museums

¹⁸ <http://herpnet.org/>.

¹⁹ <http://www.iobis.org/>.

²⁰ <http://splink.cria.org.br/>.

²¹ <http://www.gbif.net/portal/index.jsp>.

²² <http://www.biogeomancer.org/yu/>.

²³ <http://speciesanalyst.net/>.

²⁴ <http://www.anbg.gov.au/avh.html>.

²⁵ <http://www.landcareresearch.co.nz/>.

²⁶ <http://www.bgbm.fu-berlin.de/BioDivInf/Projects/ENHSIN/>.

²⁷ <http://www.cria.org.br/>.

²⁸ <http://www.gbif.org>.

collaborated to create the original MaNIS network: Bernice P. Bishop Museum, California Academy of Sciences, Colección Nacional de Mamíferos (Mexico City), Field Museum, Los Angeles County Museum of Natural History, Louisiana State University Museum of Natural Science, Michigan State University Museum, University of California Berkeley Museum of Vertebrate Zoology, Royal Ontario Museum, Texas Tech University Museum, University of Alaska Museum, University of Kansas Natural History Museum, University of Michigan Museum of Zoology, University of New Mexico Museum of Southwestern Biology, University of Puget Sound Slater Museum, University of Washington Burke Museum, and the Utah Museum of Natural History. The Principle Investigators and lead programmer of the project wish to express their sincere gratitude to the curators, staff and the support personnel at these institutions who have worked enthusiastically and collaboratively to create the MaNIS network.

We also wish to extend our sincere gratitude to the curators and staff of four additional institutions that have made major contributions to the project's collaborative georeferencing activities: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO); Sternberg Museum, Fort Hays State University; University of Kansas Natural History Museum, Division of Birds; and the University of Minnesota Bell Museum.

The MaNIS project is equally indebted to the following individuals and institutions for their programming expertise and contributions to almost all aspects of the project: PJ Schwartz (DiGIR portal developer), Dave Vieglais (DiGIR provider developer, University of Kansas Biodiversity Research Center), Reed Beaman (Yale University), Stan Blum (California Academy of Sciences), Renato Giovanni (CRIA), Australia National Botanical Garden (ANBG), Berkeley Digital Library Project (DLP), Biological Collection Access Service for Europe (BioCASE), Committee on Data for Science and Technology (CODATA), Centro de Referência em Informação Ambiental (CRIA, Brazil), Global Biodiversity Information Facility (GBIF), Taxonomic Databases Working Group (TDWG), University of Kansas Biodiversity Research Center (KUBRC).

REFERENCES

- Brook, M. de L. 2000. Why museums matter. *Trends Ecol. Evol.* 15:136-137.
- Frazier, C.K. 2002. Definition of the minimal attributes of an acceptably georeferenced specimen record. http://georef.peabody.yale.edu/tdwg/Lisbon2003/DefGeoreferencedSpecimen_2003-03-25.rtf.
- Graves, G.R. 2000. Costs and benefits of web access to museum data. *Trends Ecol. Evol.* 15:374.
- Hafner, M.S., W.L. Gannon, J. Salazar-Bravo and S.T. Alvarez-Castañeda. 1997. Mammal Collections in the Western Hemisphere: A Survey and Directory of Existing Collections. American Society of Mammalogists. <http://www.mammalsociety.org/committees/commsys/collection/collsurvey.pdf>.
- McLaren, S.B. 1988. Guidelines for usage of computer-based collection data. *J. Mammal.* 69:217-218.
- McLaren, S.B. 1999. Documentation standards for automatic data processing in mammalogy. Version 2.0. American Society of Mammalogists.
- Peterson, A.T., D.A. Vieglais, A.G. Navarro-Sigüenza and M. Silva. 2003. A global distributed biodiversity information network: Building the world museum. *Bull. Brit. Ornithol. Club* 123A:186-196.
- Vieglais D.A., E.O. Wiley, C.R. Robins and A.T. Peterson. 2000. Harnessing museum resources for the census of marine life: The FISHNET project. *Oceanography* 13:10-13.
- Wieczorek, J., Q. Guo and R.J. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *Int. J. Geogr. Inf. Sci.* 18:745-767.
- Wiley, E. and C.R. Robins. in prep. Fish of the World: FishNET as an Example Integration of Ichthyological Data. *Biodiversity Informatics*.
- Williams, S.L., M.J. Smolen and A.A. Brigida. 1979. Documentation standards for automatic data processing in mammalogy. Museum of Texas Tech University, Lubbock, TX.
- Wilson, D.E. and D.M. Reeder. 1993. *Mammal Species of the World: a taxonomic and geographic reference*. 2nd edition. Association of Systematics Collections and the Smithsonian Institution Press, Washington, D.C. <http://www.nmnh.si.edu/msw/>.