Feature Endeavour Vol. 37 No. 3
Full text provided by www.sciencedirect.com
SciVerse ScienceDirect

Computerizing natural history collections

Mary E. Sunderland*

University of California, Office for History of Science and Technology, 543 Stephens Hall #2350, Berkeley, CA 94720-2350, USA

Computers are ubiquitous in the life sciences and are associated with many of the practical and conceptual changes that characterize biology's twentieth-century transformation. Yet comparatively little has been written about how scientists use computers. Despite this relative lack of scholarly attention, the claim that computers revolutionized the life sciences by making the impossible possible is widespread, and relatively unchallenged. How did the introduction of computers into research programs shape scientific practice? The Museum of Vertebrate Zoology (MVZ) at the University of California, Berkeley provides a tractable way into this under-examined question because it is possible to follow the computerization of data in the context of long-term research programs.

Building a database for scientific collections

Today it is difficult to untangle the life sciences from the information sciences. Bioinformatics is burgeoning and interconnected databases are increasingly central to the work of biologists. These developments are often attributed to advances in molecular biology but the establishment of electronic databases for scientific collections also anchored the life sciences to computers in important ways.[1] Natural history museums captured the interest of computer programmers and applied mathematicians. Museums held elaborate, well-characterized data that seemed ready to plug into computerized databases.[2] For the most part, the interest was mutual. Natural history museums were struggling with an image problem – their research was increasingly judged to be old-fashioned. The wave of new technologies that emerged in the life sciences during the mid-twentieth century contributed to the perception that natural history was not much more than stamp collecting.[3] Those working in natural history institutions were under pressure to modernize their methods.[4] Numerical taxonomists were convinced that computers would revolutionize taxonomy in the same way that microscopy had transformed biology in the nineteenth century. While numerical taxonomy promised to make systematics more rigorous, in part by employing the analytical powers of emerging computer technologies, collections managers and curators were more intrigued by the possibilities of computerized data banking.[5] Computers promised to update natural history, first and foremost, by improving curatorial procedures; research was only a secondary and passing concern.[6] Writing in the journal *Taxon* in 1974, Stanwyn Shetler labeled these promises as myths and warned that the computer 'has a greater ability to enslave than to liberate.'[7] How did the introduction of computer databases into natural history collections shape research? Tracing activities in Berkeley's Museum of Vertebrate Zoology (MVZ) offers some interesting and perhaps unexpected answers.

Looking at the MVZ's research and curatorial practices draws attention to the historical contingencies and institutional arrangements that made computerization possible. The MVZ story also reveals the richness of natural history databases in the early twentieth century. In fact, the early electronic data processing tools were not equipped to accurately represent the interrelated multimedia content of natural history collections. For example, although natural history museums had long been accumulating photographs, drawings, field notes, and correspondence that were associated with specimens, the early databases did not provide a means to store these objects nor offer a way of representing connections between them. As a result, some specimen data was computerized while

Corresponding author: Sunderland, M.E. (Mary.Sunderland@berkeley.edu, mary.sunderland@gmail.com)
Abbreviations: ASM: American Society of Mammalogists; GRP: Grinnell Resurvey Project; MVZ: Museum of Vertebrate Zoology; NSF: National Science Foundation; SELGEM: SELf-Generated Master; TAXIR: TAXonomic Information Retrieval.
*Tel.: +1 510 642 4581.

[1] Lily E. Kay, *Who Wrote the Book of Life. A History of the Genetic Code* (Stanford: Stanford University Press, 2000); Joseph November, *Digitizing Life: The Introduction of Computers to Biology and Medicine* (Ph.D. Dissertation, Princeton University, 2006); Bruno J. Strasser. 2010. Collection, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's *Atlas of Protein Sequence and Structure*, 1954-1965. *Journal of the History of Biology* 43: 623-660; Bruno J. Strasser, 2011. The Experimenter's Museum: GenBank, Natural History, and the Moral Economy of Biomedicine. *Isis* 102: 60–90; Joseph November, *Biomedical Computing: Digitizing Life in the United States* (Baltimore: Johns Hopkins University Press, 2012); Miguel García-Sancho, *Biology, Computing, and the History of Molecular Sequencing* (Basingstoke: Palgrave Macmillan, 2012).

[2] Joel B. Hagen. 2001. The Introduction of Computers into Systematic Research in the United States during the 1960s. *Studies in the History and Philosophy of Biological and Biomedical Sciences* 32 (2): 291–314; Christine Hine, *Systematics as Cyberscience: Computers, Change, and Continuity in Science* (Cambridge: MIT Press, 2008).

Available online 7 May 2013

[3] Kristin Johnson. 2007. Natural History as Stamp Collecting: A Brief History. *Archives of Natural History* 34: 244–258.

[4] Toby Appel, Shaping Biology: The National Science Foundation and American Biological Research, 1945-1975 (Baltimore, MD: Johns Hopkins University Press, 2000); Hagen, 2001; Lynne K. Nyhart, Modern Nature: The Rise of the Biological Perspective in Germany (Chicago: University of Chicago Press, 2009); Mary E. Sunderland. 2012. Modernizing Natural History: Berkeley's Museum of Vertebrate Zoology in Transition. *Journal of the History of Biology*, doi: 10.1007/s10739-012-9339-3; Keith Vernon. 1993. Desperately Seeking Status: Evolutionary Systematics and the Taxonomists' Search for Respectability 1940-1960. *The British Journal for the History of Science* 26: 207–227.

[5] Keith Vernon. 1988. The Founding of Numerical Taxonomy. *The British Journal of the History of Science* 21: 143-159; Hagen, 2001.

[6] On research as a secondary and passing concern, see Mary Sunderland interview with James Patton, January 10, 2013, MVZ, Berkeley, CA. On the allure of computers to natural historians see David Sepkoski. 2012. Toward 'A Natural History of Data': Evolving Practices and Epistemologies of Data in Paleontology, 1900-2000. *Journal of the History of Biology* doi: 10.1007/s10739-012-9336-6 and Bruno J. Strasser. 2012. Data-driven sciences: From wonder cabinets to electronic databases. *Studies in the History and Philosophy of Biological and Biomedical Sciences* 43: 85–87.

[7] Stanwyn Shetler. 1974. Demythologizing Biological Data Banking. *Taxon* 23: 17-100, 71.

www.sciencedirect.com 0160-9327/$ – see front matter http://dx.doi.org/10.1016/j.endeavour.2013.04.001

other data remained embedded in photographs, field notes, and correspondence.

In the 1970s it wasn't necessarily practical to build computerized databases. Not only was the technology unable to capture the full scope of collections data, it also relied on mainframe computing. These complicated machines were very expensive to operate and usually required the establishment of distinct departments or centers with separate budgets and staff. As a result, there were significant barriers to their use, but many institutions pushed forward with electronic data processing initiatives. Even though the user community wasn't yet ready to fully embrace the computerized database as a research tool, the technology enabled subtle changes in the way work was done. Over time, these subtle differences cumulated to effect substantial change. It makes sense, therefore, to consider the effects of digital databases on different time scales and in different local contexts. Focusing on the efforts to digitize the MVZ's mammalogy collection during the late 1970s and early 1980s provides a window on the short-term effects of introducing electronic databases into natural history museums.

Demarcating short-term from long-term change is revealing when asking questions about how computers shaped research because it helps to unpack claims about whether or not computers were revolutionary.[8] Focusing on different time scales shows that although computerized databases initially had very little impact on day-to-day activities, at least for decades after they were built, their small influence effected substantive changes that ultimately ended up changing the character of the work. In the long-term, it is evident that computerization, more generally construed, fundamentally changed research. This large-scale change, however, was dependent on the coalescence of many different computerization efforts, including the digitization of scholarly journals, the emergence of the internet, the prominence of personal computing, as well as the computerization of databases. Looking at the MVZ's story reveals the short- and long-term costs and benefits of early electronic data processing. Although these early computer databases did not obviously impact research, the act of computerizing data contributed to a shift in how different kinds of data were valued. Because it was not possible to computerize photographs, images, and field notes, these types of data were no longer gathered or archived with the same kind of rigor. Although it is true that some investigators continued to take photographs and keep meticulous field notes, they were not obviously a part of the MVZ's database. Objects, such as photographs, and their affiliated data were located on the periphery and therefore became less visible to researchers, especially to those who might not physically visit the museum.

Now that the technology is available, the MVZ, along with other institutions, is making an effort to digitize its field notes and photographs while also imagining new ways to analyze photographs to extract ecological information. These recent developments are exciting, but they also draw attention to the potential long-term effects of excluding photographs from the earlier database. Finally, focusing on the effects of computer databases in the short term draws attention to the symbolic power of the MVZ's early computerization efforts. Computerization communicated the MVZ's commitment to engage in cutting-edge research and demonstrated its willingness to embrace new technologies – characteristics that attracted new researchers from across disciplinary and national borders.

In 1977 the MVZ submitted a proposal to the National Science Foundation (NSF) to initiate the 'computerization' of their collections. Transferring from a 'hand-written' to a 'computer organized' approach was motivated by the desire to improve curatorial practices and ultimately expand the ways in which the data could be used and disseminated. The director of the NSF's Biological Research Resources Program actively solicited the MVZ proposal because the institution had earned a reputation as being an effective early adopter of a variety of technological approaches to working with collections, such as electrophoresis. [9] The MVZ seemed like a good place for the NSF to grow its computerization efforts.[10]

Founded in 1908, the MVZ was designed as a research institution. Joseph Grinnell, the MVZ's founding director envisioned a place that could facilitate long-term studies of evolution. In fact, Grinnell predicted that the MVZ's collections would not realize their true value until 100 years had past. With the future in mind, Grinnell put in place a variety of standardized procedures to ensure that the MVZ's growing collections were scientifically valuable and well archived.[11] The result was a large database connecting objects and information deemed relevant to evolutionary questions (Figure 1).

According to Grinnell, the scientific value of each specimen depended on the information associated with it. Grinnell designed the MVZ's infrastructure to keep track of each specimen's detailed locality information in addition to any available ecological and behavioral information. Grinnell cared deeply about localities because he was interested in biogeography, speciation and subspeciation; he wanted to better understand the relationship between species formation and the kinds of geographical and ecological boundaries, such as lakes and mountains, which existed in the natural environment. This mattered to Grinnell because he was developing a theory of speciation. Although he never had a chance to write his big book on the topic, the

8 Agar, Jon. 2006. What difference did computers make? *Social Studies of Science* 6: 869-907.

9 Sunderland, 2012.

10 William Z. Lidicker and James L. Patton. 26 April 1978. 'Operational Support of the regular collection of mammals in the Museum of Vertebrate Zoology.' National Science Foundation proposal, MVZ Miscellaneous Archives, MVZ, Berkeley, California, p. 2. The MVZ's solicitation to submit a proposal was discovered during Mary Sunderland's interview with James Patton, September 17, 2012, MVZ, Berkeley, CA; Toby Appel, *Shaping Biology: The National Science Foundation and American Biological Research, 1945-1975* (Baltimore: Johns Hopkins University Press).

11 Joseph Grinnell. 1910. The Methods and Uses of a Research Museum. *The Popular Science Monthly* 77:163–169. For more on the early history of the MVZ see Elihu M. Gerson, *The American system of research: Evolutionary biology, 1890-1950* (Ph.D. Dissertation, University of Chicago, 1998); James R. Griesemer and Elihu M. Gerson. 1993. Collaboration in the Museum of Vertebrate Zoology. *Journal of the History of Biology* 26: 185-204; James R. Griesemer 1990. Modeling in the museum: On the role of remnant models in the work of Joseph Grinnell. *Biology and Philosophy* 5: 3-36; Barbara R. Stein, *On her own terms: Annie Montague Alexander and the rise of science in the American West* (Berkeley: University of California Press, 2001); Mary E. Sunderland. 2011. Teaching Natural History at the Museum of Vertebrate Zoology. *British Journal for the History of Science*, doi:10.1017/S0007087411000872.

Figure 1. Joseph Grinnell, the founding director of the Museum of Vertebrate Zoology (MVZ) at the University of California, Berkeley, organized the MVZ's database as a tool to facilitate long-term studies of evolution. Grinnell directed the MVZ from its opening in 1908 until his unexpected death in 1939. Photographer: Joseph S. Dixon. Date: 17 January 1928. MVZ photograph 6116.

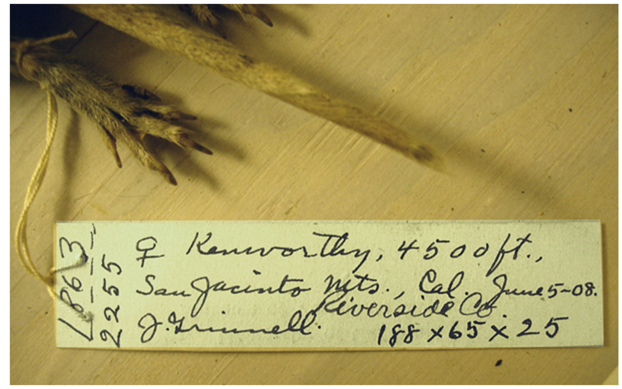

Figure 2. In the field, a tag is attached to each specimen. The tag lists the specimen's sex, measurements, locality, date, as well as the collector's name and field number. After the specimen has been entered into the MVZ's catalog, this number is also added to the tag. To allow for taxonomic revisions, the genus and species name are written in pencil. Photographer: Karen Klitz, 2011.

core of his convictions can be gleaned from his publications, field notes, correspondence, and importantly, his lecture notes.[12]

Grinnell saw that habitats were dynamic in two ways. First, they were being altered by human use and second, they varied spatially and shifted temporally. This observation is not surprising, especially considering the rapid urban and farming developments that were happening across California; habitats were dramatically altered from one day to the next. The changing environment was forcing animals to move. When an animal found itself in a new environment it would have to adapt, keep moving, or perish. Grinnell set out to document this process. Human developments aside, California was and is an extraordinarily diverse landscape. There are mountains and valleys, rivers and oceans, deserts and plains. Grinnell was interested to document the inter- and intra-species diversity within, and on the edges of, these diverse places. So, it mattered very much if a mouse was found on black lava rock in the desert or on the adjacent white sand. Just like it mattered if a rabbit was found under a juniper tree or next to a barrel cactus.

The MVZ was organized to keep track of all of these details. A detailed tag was created and tied to each specimen in the field (Figure 2). Grinnell insisted that collectors keep scientifically credible field notes that were also theoretically interesting. It was especially important to keep track of maps, weather conditions, and general ecology. Grinnell established careful instructions for taking proper field notes. Students in Grinnell's Natural History of the Vertebrates class learned this skill during field excursions with their teacher. Graduates in the MVZ were provided with the same detailed instructions and mentoring to ensure that their field notes were not just useful for their immediate research project, but also in the long-term. First and foremost, it was of utmost importance to record observations and ideas while in the field. Grinnell did not trust his memory, or those of his students. To ensure that field notes would last the test of time, Grinnell insisted on certain materials to guarantee a permanent record, such as India ink and archival (100% rag bond) paper. Next, it was necessary to separate general field notes and species accounts. General field notes were for keeping track of things like the weather, time, and miscellaneous observations whereas species accounts involved detailing almost everything about a particular animal that was being observed.[13] Grinnell emphasized, 'Every sort of fact definitely observed should be recorded; and observations even of the very same nature should be repeated again and again, as opportunity permits, for each species.' The scientific value of each specimen was determined, in part, by its affiliated field notes.[14]

The MVZ was therefore organized to link together specimens, and their associated ecological and geographic data through the specimen label and collectors' field catalogs and journal entries. In addition, collectors were encouraged to take photographs that could compliment and illustrate the observations laid out in field notes. Photographs were also cataloged. Finally, Grinnell kept careful track of the detailed correspondence that took place within the MVZ as well as external communications. A card catalog kept track of which letters discussed which species. Card catalogs and ledgers kept track of and connected the information that was on specimen tags, in photographs, in letters, and in field notes. It was a very

[12] Sunderland, 2011.

[13] Interestingly, Grinnell himself never wrote a species account. For more on Grinnell's field notes see John D. Perrine and James L. Patton. 2011. 'Letters to the Future: Field Notes and the Grinnell Resurvey Project,' in *Field Notes on Science and Nature* (M. Canfield ed.), Harvard, 211-250 and Cathryn Carson. 2007. Writing, writing, writing: The natural field journal as literary text. *Townsend Newsletter*, 6-8. For more on instructions for field collectors see Jim Endersby, *Imperial Nature: Joseph Hooker and the Practices of Victorian Science* (Chicago: University of Chicago Press, 2008) and Kristin Johnson, *Ordering Life: Karl Jordan and the Naturalist Tradition* (Baltimore: Johns Hopkins University Press, 2012).

[14] Joseph Grinnell, 'Suggestions for Field Notes,' Joseph and Hilda Wood Grinnell Papers, Bancroft Library, Berkeley (subsequently JHGP), Box 9, Folder 8; Sunderland, 2011.

effective database that organized a diversity of materials and information.

Making catalogs digital

Card or ledger catalogs continued to be a widespread technology in natural history museums throughout the twentieth century.[15] Hand-written tags, notes, and catalogs organized vast collections. In the 1960s, some museums began to introduce a variety of automation technologies. The Smithsonian National Museum of Natural History participated in the development of SELGEM, which stood for SELf-Generated Master. Data were entered into the SELGEM platform with numeric tags and special characters. SELGEM was able to process the data in a variety of ways, including creating different forms and tables. In 1975, Jerome G. Rozen, Jr., Deputy Director for Research at The American Museum of Natural History sent a memo to the chairpersons of curatorial departments across North America, alerting them to these computer services that were available through the Smithsonian. The memo included a detailed description of the computer services.[16] A few organizations experimented with SELGEM, but it was not widely adopted. Disk storage and processing costs were expensive on mainframe computer systems. Although SELGEM was free and included a number of appealing features, 'its virtues' were 'off-set by its unwieldiness and inefficiency.'[17] The MVZ's curator of mammalogy, James Patton, noted that the University of California was using SELGEM for administrative purposes and predicted that the MVZ wouldn't be making use of the system.[18]

Natural history museums were not quick to embrace electronic data processing because the added value of computerizing natural history data was not entirely evident. Many natural history museums had scarce resources that were needed to simply maintain existing collections. The uncertain future of collections motivated a variety of meetings and international conferences – it was time for collections to find a common identity. In the early 1970s curators from a diversity of natural history institutions gathered to discuss the future of natural history collections and identify commonalities across collections. What were the shared needs and practices affiliated with collections? The need to computerize records was one of these shared needs, but the concern did not rise to the top of the list.[19] Although incorporating computers into curatorial practices did not seem urgent, the mammalogy community was well poised to start thinking about the possibilities of doing so.

Why mammals? In addition to being one of the smaller taxonomic groups, the North American mammalogy community was tightly linked. Camaraderie among North American mammalogists was boosted by their shared academic lineage. Many of the mammalogy curators in American institutions and leaders within the American Society of Mammologists (ASM) traced their academic roots back to Grinnell.[20] And many continued to use the same 'Grinnellian' methods of taking field notes, collecting locality data, and writing specimen tags. The MVZ's curatorial procedures were widespread, at least throughout the mammalogy community, thanks to the many generations of Grinnell students who populated museums and held curatorial positions. In effect, a lot of the data from the mammalogy community had already been standardized and were therefore more easily transferred into electronic form.

In 1972, the ASM established the Committee on Information Retrieval to organize thinking about building a shared computerized information retrieval network for mammal collections. Together, the Committee suggested that the ASM find a way to make SELGEM work for mammal collections and began crafting a proposal to the NSF to request funds for a pilot project. The ASM hoped to build a national information retrieval network, linking together the many mammal collections.[21] Motivations to build the information retrieval network were practical. Many curators were overwhelmed by the clerical activities that were involved with responding to requests about specimen data. Hours could be spent on a single request that inquired as to whether a certain species was housed in an institution's collection. SELGEM promised to make these kinds of queries infinitely easier.

The early interest in electronic data processing was motivated by collections-management concerns rather than research possibilities. With practical concerns at the forefront, not all were enthusiastic. Curators who had experience with computerizing wrote to share concerns and advice. Stephen Humphreys, assistant curator in mammalogy at the Florida State Museum at the University of Florida, warned, 'We have found the cost to be far higher than is widely understood.' He outlined 'hidden subsidies' that might result in unreasonably low estimates and hoped that his advice would help to avoid 'the nightmare' that he 'foresaw in the original ASM proposal.' Initial cost estimates often failed to account for: machine costs, including extramural specimen data, and the time invested by curatorial technicians. 'This is a big job, and someone needs to do nothing else but supervise it. If the supervisor is only titular, the job will not be accomplished.'

15 Goode, George Brown. *The origins of natural science in America: The essays of George Brown Goode*, Sally Gregory Kohlstedt ed. (Washington, DC: Smithsonian Institution Press, 1991); JoAnne Yates, *Control Through Communication: The Rise of System in American Management* (Baltimore: Johns Hopkins University Press, 1989).

16 Jerome G. Rozen, Jr. to Chairpersons of Curatorial Departments, 6 February 1975, Smithsonian Institution Archives, Record Unit 7357, American Society of Mammalogists (ASM), Records, 1919-1993 and undated (subsequently ASM records), Box 133, Folder 1, Washington, DC.

17 Committee on Information Retrieval of the ASM (Koeppl, J.W., Chairperson), 1982, Computerized Information Retrieval in Mammal Collections of North America by, ASM records, Box 95, Folder 7, Washington, DC.

18 James L. Patton to Sydney Anderson, 4 May 1976, ASM records, Box 133, Folder 1, Washington, DC.

19 Humphrey, Philip S. 1972. A New Organization: The Association of Systematics Collections. *Curator: The Museum Journal* 15: 32–33; Sunderland interview with Patton, 17 September 2012.

20 J. Knox Jones Jr, 1991. 'Genealogy of twentieth-century systematic mammalogists in North America: the descendants of Joseph Grinnell,' in Michael A. Mares and David J. Schmidly (eds.), *Latin American Mammalogy: History, Biodiversity, and Conservation*. Oklahoma: University of Oklahoma Press, 1991, pp. 48–56; J.O. Whitaker, 1994, 'Academic propinquity: III. The Joseph Grinnell/E.R. Hall group (Berkeley and Kansas),' in Elmer C. Birney and Jerry R. Choate (eds.), *Seventy-Five Years of Mammalogy, 1919–1994*, Provo: American Society of Mammalogists, pp. 129–134.

21 'Report and Recommendations: Advisory Committee for Systematic Resources in Mammalogy.' 1974, MVZ Miscellaneous Archives, MVZ, Berkeley, California, pp. 24-28.

Humphreys hoped that his experience might help the ASM to 'propose a realistic job.' Although Humphreys asserted the clear 'value of computerization' he emphasized that it is only possible with adequate resources. His letter drew attention to the cost difference between entering full data versus partial data. The Florida State Museum opted for the costlier full data, 'convinced that a partial-data system will provide little more retrieval power than the currently conventional letter of inquiry.' Humphrey emphasized the need to generate a retrieval system that would also help with research, 'not just tell what is inside our cabinets (we can learn that far more cheaply by looking).'[22]

How might electronic data processing help with research? This was also a concern of Dexter Hickey, Associate Director of Research at the Institute of Ecology. Hickey wondered how a study on data retrieval might benefit research in evolution and/or applied work in wildlife studies, disease prevention and pest control. Hickey recommended that the ASM's proposal needed to directly tie the data retrieval work to research. 'For example, you might postulate from published records, a range for an endangered species, then find out to what extent unpublished museum records confirm or modify the predication. Such an exercise would help show how much new knowledge can be obtained by compilation of old records.'[23]

Ultimately, for not entirely clear reasons, the ASM's proposal to the NSF was not successful. Curators were encouraged to pursue computerization efforts within their own institutions. Having read through many versions of the proposal, members of the Committee on Information Retrieval were well versed in the language of computer databases (at least better than the average mammalogist). James L. Patton, curator of mammalogy at the MVZ was one of the committee members who went on to submit a computerization proposal through his home institution. Together with fellow MVZ mammalogy curator, William Z. Lidicker, Patton submitted a proposal to the NSF, requesting $153,059 to develop a computer-based catalog system for the MVZ's mammal collection. Writing in 1978, the proposal's authors, considered electronic data processing 'the most useful recent development to be incorporated as a managerial system in museum collections.'[24] There was some evidence to support their claim. Although concerns about cost-effectiveness remained, the process was becoming cheaper and more widespread. The authors pointed to at least ten other collections that were developing electronic-based programs. There was special motivation to focus on the MVZ's mammal collection. The larger mammalogy community was committed to establishing electronic processing procedures and had already established guidelines for required data fields for mammal collections and minimal standards for the data contained within each.

Rather than SELGEM, Taxir was the MVZ's program of choice. Their botanist colleague, Tom Duncan, was already familiar with the program and it was possible to make it available through the local university computer system. The MVZ opted for Taxir for the sake of convenience, but also because the NSF encouraged them to do so. In the late 1960s, the NSF had funded the development of Taxir as a computerized storage and retrieval system to aid the curation of systemic collections. Fittingly, Taxir stands for TAXonomic Information Retrieval. Led by the botanist, David J. Rogers, Taxir was designed as a tool to meet the needs of curators. Much like Excel today, Taxir was a flat-field database program. It used common fields to store specimen data, which enabled efficient searching. It was possible to generate reports in response to queries, such as 'what taxa occur in area A' or 'what localities do we have for taxon Z.'[25] One of Taxir's developers wrote a primer to make Taxir available to everyone who needed to retrieve select data from a large data set, regardless of the nature of the data. The hope was that Taxir would be adopted for a range of administrative purposes. Taxir's developers were convinced that it could be used by 'any intelligent motivated person' even those without scientific education.[26] NSF was invested in the success of Taxir and saw the MVZ as a model institution that would allow them to test its efficacy and also convert the system to function on an IBM computer system.

Taxir has been developed at the University of Colorado and was being used by the herbarium at the University of Michigan on an Amdahl mainframe system. By the late 1970s, many institutions were adopting IBM systems and so the NSF likely saw the need to transfer the Taxir program. Along with the MVZ, the NSF invited the mammalogy curators at the University of Michigan Museum of Zoology to submit a computerization proposal. In many ways, the two institutions mirrored each other. Both of the mammalogy curators, Emmet T. Hooper and Philip Myers, were MVZ alumni. The cataloging systems and practices at the two institutions shared much in common. Looking at both how both institutions adopted Taxir would allow the NSF to test the program. Furthermore, the two museums envisioned a collaboration built around their electronic efforts.

In 1978 there were 154,000 mammal specimens in the MVZ. For each specimen, the authors proposed to capture twenty-four data fields. In addition to the categories that had been approved by the ASM's Committee on Information Retrieval, the MVZ identified an additional set needed to facilitate their curatorial practices. The mandatory categories included: institutional acronym; catalog number; genus; species; type(s) of preservation; sex; date collected; continent or country; state or province; county, parish, district, department, major island group; ocean; sea; and bay, inlet, strait, estuary, gulf, or channel. The MVZ's additional fields included: accession number; order; family; subspecies; collector's name; collector's catalog number; preparator's name; preparator's catalog number; specific locality; township, range, and section; elevation; and remarks. The new MVZ fields reflect Grinnell's understanding of what makes a specimen value: its affiliated information. The collector's identity matters

[22] Stephen Humphrey to Hugh Genoways, 28 October 1975, ASM records, Box 133, Folder 1, Washington, DC.

[23] Dexter Hickey to Jerry Choate, 8 May 1975, ASM records, Box 133, Folder 1, Washington, DC.

[24] Lidicker and Patton, 1977, p. 1.

[25] Patton to Sunderland, 19 September 2012.

[26] Brill, RC. The Taxir Primer. Occasional Paper, no. 1. Institute of Arctic and Alpine Research, University of Colorado. 1971, pp. 5-6.

because it ties the specimen to field notes, just as the elevation at which the specimen was collected matters because it is contains critical locality information. However, the reasons for these additional fields are implicitly tied to the MVZ's curatorial practices. For example, the proposal indicates that these specific fields were added because they already exist in the card catalog. Stated plainly, 'We are not interested, frankly, in building an EDP [electronic data processing] file that provides us with less available information than our current procedures.'[27] The MVZ proposed to thoroughly capture all data listed on the existing specimen catalog cards. What was not captured? Field notes, photographs, and correspondence; the implications of de-emphasizing and/or removing these data will be discussed in the next section. Of course, these data remained in the MVZ and continued to grow, but was not included as a core part of the curatorial practice. From Grinnell's day to the very recent past, data resident in field notes and in some cases photographs, were only secondarily accessible through the collector's name and date of collection.

By the end of the three-year grant, the MVZ's mammalogy collection had been almost entirely computerized. By 1981, only a few thousand specimens still needed to be retroactively captured and so the MVZ went looking to the NSF for additional support. In the second proposal they requested money to hire personnel, including a computer programmer – someone who was able to write local code (data entry, etc.) and editing protocols. Although UC Berkeley was willing to support the Taxir maintenance on the IBM system, there were no personnel available to conduct further developments. Second, they requested money to pay for continued disk rental and access to data banks. Thirdly, they requested money to develop new data banks for the frozen tissue and anatomical collections. This third request was the main component of the second proposal because they realized that these collections had different data requirements – they were meant to be used. For example, a piece of tissue would be removed from a frozen tissue specimen and then used up during the relevant molecular analyses.[28] Finally, they requested money to support the costs affiliated with providing user access. Generating reports to users' queries cost money and the MVZ was not prepared to absorb the cost. Most user requests now required them to check the data bank and print out query results, which cost money.[29] This second NSF grant was awarded to the MVZ to support further digitization efforts centered on the mammal collection. It should be noted, however, that Taxir was just the first step of many taken to computerize the MVZ's collections. And, although cost reduction was one of the motivations for computerizing, the MVZ would continue to invest substantial resources into digitization, indeed, investments continue today. During the 1980s, the MVZ submitted multiple NSF grants to support the computerization of its ornithology and herpetology collections, and in the 1990s, more funding was acquired to support the transfer of its database from Taxir to a relational database that could be made available online. VertNet (vertnet.org), the MVZ's most recent data-centered research endeavor, involves an international effort to bring together distributed databases and ultimately make specimen data from a diversity of institutions interoperable, mappable, and accessible.

Transforming research

Before jumping ahead to VertNet and considering the new research possibilities it enables, this section asks whether the MVZ's Taxir-based mammal database allowed users to engage in different kinds of research. Were new queries possible? Were results of queries different than when they were conducted manually? Tracing the MVZ's research on pocket gophers begins to answer these questions and suggests that the computerized database had a gradual, cumulative effect on the kinds of research that were possible. Initially this change looks to be a difference in scale rather than a fundamental conceptual shift. This does not mean that the research didn't change over time. It changed significantly. More obvious change, however, resulted from new molecular approaches, such as electrophoresis and DNA sequencing technologies, rather than because the collections database was digital. Looking more closely, however, suggests that computerizing records instigated a subtler shift in how researchers placed scientific value on the different kinds of collections and their associated data.

The pocket gopher story is a window on changing research practices during the digital transition because the research project's core can be traced back to Grinnell's studies in the 1920s and followed through to pre-computerization work by Patton in the 1970s, and later in the 1980s and 1990s after digital records and retrieval were well established. It's important to note that although Patton's work in the 1970s was prior to the digitization of the mammal collection, it was not conducted without the aid of computers. In the early 1970s, Patton began using computer applications for multivariate statistical analyses of morphological variation and for building allozyme trees. The development of these kinds of analytical programs, however, was not dependent on the establishment of computer databases. It was possible to do one without the other. After two decades of studying pocket gophers, Patton and his colleague and former graduate student, Margaret (Peg) Smith, published a treatise of their cumulative studies, 'The Evolutionary Dynamics of the Pocket Gopher *Thomomys bottae*, with Emphasis on California Publications.' The dedication is to 'Carol Porter Patton, an intrepid trapper and field companion' – Patton's wife, who accompanied him on many of his field excursions, and to 'Joseph Grinnell, in whose memory this project has been completed.'[30]

Grinnell wanted to learn more about the role that geography played in speciation processes. In the MVZ' s

[27] Lidicker and Patton, 1977, p. 4.

[28] As it turns out, they were not actually able to develop a new way to track specimen use with the Taxir system. Doing so was not technically feasible until the MVZ moved to a relational database program in the 1990s.

[29] 'Support for the EDP (Electronic Data Processing) Program of the Mammal Collection of the Museum of Vertebrate Zoology, University of California, Berkeley, Summary of Accomplishments Under NSF Grant DEB 78-07110,' MVZ Miscellaneous Archives, MVZ, Berkeley, California.

[30] James L. Patton and Margaret F. Smith. 1990. The Evolutionary Dynamics of the Pocket Gopher, Thomomys bottae, with Emphasis on California Populations. *University of California Publications, Zoology* 123: 1–161.

Figure 3. In the MVZ, the mammal collection is stored on trays and contained in large gray metal cases. The specimens (skins and skulls) are arranged alphabetically by family, genus, species, and subspecies, then by geographical area. Photographer: Karen Klitz, 2011.

early years, very little was known about the mechanisms underlying species formation and Grinnell hypothesized a critical environmental role. He predicted that each species inhabited a specific niche, its immediate surroundings, and that the conditions of the niche were determined by a variety of factors including humidity, elevation, temperature, and vegetation. Each species was only able to occupy one niche. Grinnell's niche concept allowed him to see the environment from a taxonomical perspective because niches could be grouped in different ways.[31] Grinnell was especially interested to locate animals on the edge of their established niche, or in a new niche. Faced with a new environment, these animals would need to adapt and possibly diverge to form a new species. Focusing on niches revealed animals that seemed out of place. Pocket gophers were a particularly interesting case because even in slightly different environments their appearance changed significantly. Grinnell hypothesized that these different looking animals were different subspecies and wondered how and why there were so many different subspecies. By carefully describing each subspecies, including detailed notes on habitats and behaviors, Grinnell gathered important data that could be used to study speciation and understand evolutionary processes.[32]

Today, there are over 21,000 pocket gophers in the MVZ (Figure 3). The collection dates back to the MVZ's earliest days. Following the pocket gopher specimens from the field into the museum's collections and onto the pages of research publications draws attention to the role of computers in this process and raises questions about the effects of the MVZ's early digitization efforts. Obviously, there were no computers involved in Grinnell's day, but there was a great amount of labor expended to carefully archive a diversity of materials and information, in particular, specimens, photographs, field notes, correspondence, and locality information. Index cards were standardized with a list of fields that had to be filled in at the time a new specimen was accessioned into the museum. And two secondary catalogs were developed to facilitate the retrieval of relevant information.

In total, the mammal collection had three catalogs: the specimen catalog consisted of a sequential list of each new mammal cataloged into the museum; the taxonomic reference catalog was organized taxonomically with each index card including the name of a species and a list of each specimen that belonged in the category; and the geographic reference catalog, which was organized by locality. Each locality index card included the name of a particular locality and a list of every MVZ specimen that had been found there. As a result, when a new mammal came into the museum it had to be entered into three separate catalogs – first a new card was made for the specimen catalog, then the specimen was entered into both the taxonomic reference catalog, and the geographic reference catalog. This organization was meant to aid the user. For example, if you were wondering if the MVZ collection had any pocket gophers, *Thomomys bottae*, from the Coachella Valley, you would go to the geographic reference catalog, look up Coachella Valley and start scanning the index cards for *Thomomys bottae*. Alternatively, if you wanted to know how many pocket gophers were in the MVZ's collection, you would go to the taxonomic reference catalog, turn to the card(s) devoted to *Thomomys bottae*, and count all of the listed specimens.[33]

Until the 1950s, species were most often identified by their physical characteristics. *Thomomys bottae* were categorized as such based on an analysis of their morphology such as their body size, skull shape, and pelage color. It was not until the 1960s that it became more common for scientists to look inside cells to uncover cryptic variation and discern evolutionary patterns and relationships. As a graduate student at the University of Arizona, Patton became interested in studying the chromosomes that could be extracted from cells. During his graduate work, Patton developed a technique to isolate mammalian chromosomes, something that had previously been very challenging and expensive.[34] The scientific community's interest in karyology (the study of chromosomes) grew after the discovery of DNA. When Patton applied his technique to study the chromosomes of pocket gophers, he noticed that pocket gophers in nearby populations often had extremely diverse karyotypes. Because different species have different looking chromosomes, Patton was able to carefully study hybrids and ask questions about population dynamics and species formation.[35]

Patton built on Grinnell's earlier observations and the way that he initially accessed the data was strikingly similar to how Grinnell might have worked. Smith, who

[31] James R. Griesemer, 'Niche: Historical Perspectives.' In: Evelyn Fox Keller and Elisabeth A. Lloyd, ed. *Keywords in Evolutionary Biology* (Cambridge: Harvard University Press, 1992), pp. 231–240.

[32] Joseph Grinnell. 1927. Geography and evolution in the pocket gophers of California. *Smithsonian Institute Annual Report* 2894: 335–343.

[33] Patton to Sunderland, 17 September 2012. For more on how specimens became a part of the MVZ's collection in the early twentieth century see Mary E. Sunderland, Karen Klitz and Kristine Yoshihara. 2012. Doing Natural History. *BioScience* 62: 824–829.

[34] James L. Patton. 1967. Chromosome Studies of Certain Pocket Mice, Genus Perognathus (Rodentia: Heteromyidae). *Journal of Mammalogy* 48: 27–37.

[35] James L. Patton. 1972. Patterns of geographic variation in karyotype in the pocket gopher, *Thomomys bottae*. *Evolution* 26: 574–586.

worked on the gopher project from the 1970s through the 1990s, recollects that she continued to access specimens through the card catalog. Although specimen records were being entered into a variety of computer databases, many museums were paranoid about doing away with their old catalogs. Computer databases were not seen as a replacement for specimen catalogs, instead they were seen as a storage and retrieval tool. The MVZ, for example, continued to maintain its card catalog until 2003, and continues to print hard copies of its electronic database as back-up records. The decision to stop keeping a hand-written catalog was motivated by a variety of reasons, but initiated largely by the retirement of Patton. Although Patton continues to emphasize the necessity of keeping a hard copy, upon his retirement he reasoned that it should, in theory, be possible to simply print out a ledger generated from the computer database. The process of printing out ledgers from the database, however, proved to be a challenging task that took many years to realize.[36] The MVZ, along with other museums, is still considering how to best maintain hard copies of computer-based data. Throughout this transition period, many museum users continued to work with the card catalogs.

The secondary catalogs (the taxonomic and geographic reference catalogs) were no longer needed. When the MVZ first built the Taxir database, Patton started printing out locality or taxon lists to bring with him into the field-lists that would have taken far too long to compile by hand. For example, before heading out to the Mohave, Patton would query the database for all of the pocket gopher records in the Mohave region, with specific localities organized by county. These lists helped to orient Patton in the field. He quickly realized that although Taxir was brought into the MVZ to widen the collection's use, he would be both a major user and beneficiary of the new database. Not only did Taxir help with research, it also helped with curation because entering data into the database ultimately improved the quality of the data. For example, after printing out the list of pocket gophers and their localities, Patton might notice that multiple subspecies are listed as occurring in the same locality. Since this is highly unusual, it's likely a mistake and would prompt Patton to look up those particular specimens, check their identities and modify the catalog entries accordingly.

Taxir centered on the specimens. There was a variety of carefully chosen fields to document each specimen's characteristics. In contrast, Taxir did not keep track of information regarding field notes or photographs. If you wanted to find something in the field notes, you would have to visit the MVZ's library and look through the available bound volumes of notes. Periodically, field notes were bound in a systematic way that mostly met the original standards for field notes, as laid out by Grinnell. But they were archived much less systematically. Students, faculty, and staff were implicitly expected to deposit their notes in the library, but the process to ensure this procedure was not clear. There is no database, for example, to show the authors and dates of field notes and no records that track graduate students and their notes. As a result, there are many gaps in the field note collection. The same is true for the photograph collection. Sometimes photographs are attached to the pages of field notes and sometimes not, but, for a long time, there was no formal process to track them through a database. The main reason for this was the lack of appropriate tracking technologies. These ramifications are explored further in the following section. [37]

Changing the way data are stored and accessed affects the way the data are used and valued. This statement is made evident by the way most researchers consult the scholarship in their field today. An article is much more likely to be read if it is easily available online. Furthermore, a highly cited article is more likely to be cited in the future (whether the article is actually read is another issue altogether). Entering the data from the specimen index cards into an electronic database made it accessible in a way that photographs and field notes were not, especially as the data became available online in the 1990s.

What is the scientific value of photographs taken in the field? There is little doubt of their value during Grinnell's time and today photo retake projects are employing photogrammetry methods to perform statistical comparisons of vegetation differences across historic and recent photos. Until the 1940s, taking photographs in the field was a significant expense of labor and resources. Taking pictures involved carrying large format cameras, which used heavy, glass negative plates and required a tripod. In addition to their weight, these items were extremely fragile, making them very cumbersome to take to remote field locations. Despite this obvious expense, Grinnell insisted on photographs and kept a careful database to keep track of the collection. Well known for his conservative tendencies, Grinnell was careful to allot resources only for meaningful scientific pursuits. He must have thought that photographs contained important data (Figure 4).

What kinds of data might be embedded in the photographs? What kinds of pictures were people taking? Scanning through the many photographs that were taken before Grinnell's unexpected death in 1939, it is possible to categorize the pictures by general subject matter. As expected, there are many landscape pictures that show the environment conditions of field sites. There are close-up photographs that show detailed habitat information, such as possible food sources or nests and beaver dams, as well as pictures that were taken from a distance that capture the landscapes general features, or show the location of a campsite. There is also a surprising number of photographs of living animals – over a thousand. These pictures show animal behavior, such as a coyote inflicted with porcupine quills or bears playing. In addition there are many photographs of animals on their way to becoming and after becoming museum specimens. These pictures capture the preparatory process and draw attention to the defining features of a specimen (Figure 5). Close-up photographs also document the immediate surroundings of a trap. Although there are comparatively fewer pictures of people, they do exist. Photographs document field parties,

[36] Mary Sunderland interview with James Patton, January 10, 2013, MVZ, Berkeley, CA.

[37] Procedures to archive photos digitally are actively being developed in the MVZ through the database system that they currently use, Arctos. The technological and conceptual developments that led to this possibility are beyond the scope of this paper.

Figure 4. The MVZ's early image catalog includes a photograph depicting the proper camera arrangement for photographing specimens. Photographer: Tracy I. Storer. Date: 30 August 1915. MVZ photograph 2042.

Figure 5. Taken in Fresno, California in 1916, this photograph shows (from left to right) Halsted White, Harry Swarth and Joseph Dixon preparing specimens in the field. In the catalog, the subject was listed 'Swarth, Dixon, and White at work.' Photographer: Joseph S. Dixon. Date: 21 August 1916. MVZ photograph 2227.

faculty, staff, and students at work in the field. Local culture is also documented, perhaps to show the historical records of the localities where people were collecting specimens. These photographs that were taken in the field during the Grinnell era sometimes appeared in scientific publications and other times were used for teaching purposes.

Although people in the MVZ continued to take photographs after Grinnell's death, there was an obvious change in the maintenance of the collection. Photographs were cataloged more periodically through the 1950s and 1960s, but by the 1970s the card catalog was no longer updated. This didn't mean that photography stopped, just that photographs stopped being a systematic part of the collection in the way Grinnell designed. It is possible to find photos by skimming through field notes or by contacting different collectors. Also, after the 1990s, when desktop programs like FileMaker Pro became widely used, some later photographs were accessioned into a separate database of special field excursions, such as those that involved the participation of multiple curators, but new photos were not accessioned systematically. Considering these changes regarding how photographs were treated opens up questions about the relationship between data-keeping and research practices. Photographs contain embedded data that shape the way researchers conceptualize the relationship between the collections and nature. Shifting the status of photographs from central to ancillary contributed to a larger shift that made it possible to envision specimens as isolated data, stored in a computer, rather than as contextualized data embedded in a particular natural environment, depicted in a specific photograph.

Why computerize?

Many, but not all, agreed that doing away with at least some of the hand-written index cards was a move in the right direction. Some argued that computerization had the potential to benefit all collections, while others were adamant that it should be actively avoided, especially by small collections. These diverse perspectives are evident in the ASM's 1982 report on computerization. As the development of electronic natural history collection catalogs, like the MVZ's, became more prevalent, the ASM was motivated to investigate how computerization was occurring, 'sometimes with disastrous results,' in North American mammal collections. The report identified a list of 'myths and misconceptions' about the benefits of computerization but ultimately argued that 'most collections could benefit from computerization' even small collections.[38]

The reasons for computerizing were rooted in four basic practical concerns: growth, mobility, habits, and cost. With respect to growth, the argument was that small collections should anticipate becoming large collections, especially because one of the most daunting tasks of computerizing is capturing retrospective data; if you're small, it's best to start now. Regarding mobility, if a small collection does not grow, it's likely to be incorporated into a larger collection and computerization will help to facilitate the inevitable migrations. 'Collections, like people, need to establish good habits.' The argument is that computerizing data forces a greater attention to detail, therefore the computerization process 'enhance[s] the quality and usefulness of a collection.' Finally, cost – the thought was that price is steadily decreasing. In fact, 'most people who in the past could afford a typewrite can now afford a micro-computer.'[39] Not everyone agreed about the value of computerizing

[38] Committee on Information Retrieval of the ASM, 1982, p. 1 and 11.
[39] *Ibid*., p. 12.

small collections. Phil Myers, curator of mammalogy at the University of Michigan's Museum of Zoology asserted that a curator of a small mammal collection (under 20,000 specimens) would be 'a fool to even consider computerizing.'[40]

Computerization was fueled by the promises of data banking; however, even by the late 1970s there was already a need to 'demythologize' these promises.[41] In his letter to J.W. Koeppl, criticizing the ASM report's enthusiastic endorsement of computerizing small collections, Myers went on to predict 'Maybe in five years the micro hardware and software, plus the general level of computer literacy, will intersect at a point where the average mammalogist (who, let's face it, doesn't know CP/M [Control Program/Monitor – a popular operating system for early personal computers] from ICBM [Intercontinental Ballistic Missile], either of which can wreck your life in about equal time) can spend a few days and a few bucks and have a workable system. But that day isn't here yet; it's really only people like you (who are computer literate) or me (who, because of the size of their collection and the resources of their community, can get help) who truly benefit from an investment in computerization.' Myers candidly mused '"But it's easy,' you say. Sure it's easy, if you have some idea of what you're doing. But your average curator won't, and he will resent the unexpected time he will have to put into routines he probably thought were 'off the shelf' when he bought them." Myers went on to predict that the curator would 'spend hours talking with computer salesmen, software jocks, etc. – much of the time in utter frustration.' Myers identified 'the current Catch 22 of the business' – it's not possible to determine if software really works for you until you buy it, install it, and spend a few weeks trying to make it do want you want and if you then discover that it doesn't quite work the way you thought, it's too late.[42] It's important to note that many of Myers' concerns were alleviated shortly after he made them with the introduction of the PC. The PC ushered in a new computing era that dramatically altered the way people work, in part, by reducing the barriers that Myers described.

Even though Myers was worried about small collections, he was confident that computerization would lead to benefits for institutions like his, and the MVZ. What were these benefits? Accessibility? Initially computerizing collections altered access is subtle ways. It became possible to obtain data with relatively little effort that previously would have required substantial resources. Before electronic databases it would have been unreasonable to ask a curator to provide you with a list of all of the pocket gophers in Fresno County. Obtaining the answer would require a visit to the collection to look at pocket gophers. In contrast, digital databases made this a trivial question. Answering it involved performing a simple database query and generating a computer print out. But for the most part, the user community remained the same – people that were conducting some kind of collections-based research. Accessibility did not change fundamentally until the advent of web-based technologies in the 1990s.

The creation of digital databases was also justified by practical concerns related to streamlining accession processes and increasing efficiency – saving both time and money. However, the MVZ, did not do away with index cards. In fact, they continued to upkeep the hand-written card catalog until 2003 when Patton was confident in the security of electronic databases, but as noted earlier, he hedged his bet by printing out an archival ledger of specimen data from the electronic file. Before 2003, index cards were filled out first and then the same information was entered into the electronic database. Although the two secondary catalog were no longer maintained, one is led to wonder how much time was actually saved, especially considering the amount of time if took to train curatorial assistants to enter electronic data in addition to the time that it took computer specialists to maintain the Taxir system. Patton explains that the MVZ's early database initiative was not about saving time, even though it was eventually quicker to input specimen data once, electronically, rather than writing the information three times into three separate catalogs. The core motivation was to minimize errors since recopying information multiple times increases the risk of introducing mistakes into the database.[43]

Most notably, it was expected that developing an electronic catalog would allow the collections to be used in new ways. Comparing Grinnell's attention to photographs and field notes with their lack of representation in the electronic database subtly shaped the ways in which researchers interacted with information. Although not obviously significant at first, these shifts in attention cumulated to impact perceptions of what counts as scientifically relevant data. The epistemic status of photographs, correspondence, and field notes was demoted over the long-term, in part by their initial lack of representation in the database. Additional diminishing factors include a more general shift away from the use of photographs in scientific publications. A similar trend, for example, has been documented in paleontology.[44] This trend was also spurred by changing camera technologies that made it simpler for individuals to take their own personal photographs, some of which were attached to the pages of field notes. At the same time, there was less impetus to document the collecting activities of individual investigators. After Grinnell's era it became increasingly common for a faculty member to organize field expeditions for his or her personal research program, rather than embarking on museum-wide excursions.

It's important to emphasize that until very recently, database technologies have not been equipped to deal with photograph's and field note's rich data. It's only very recently that the scientific relevance of these sources has been brought into focus. The MVZ's Grinnell Resurvey Project (GRP), funded by the NSF in 2007, highlights the research opportunities enabled by Grinnell's full database. The GRP involves returning to areas that had been

[40] Phil Myers to J.W. Koeppl, 25 July 1983, ASM records, Box 95, Folder 7, Washington, DC.
[41] Shetler, 1974.
[42] Myers to Koeppl.
[43] Patton to Sunderland, 19 September 2012.
[44] Sepkoski, 2012.

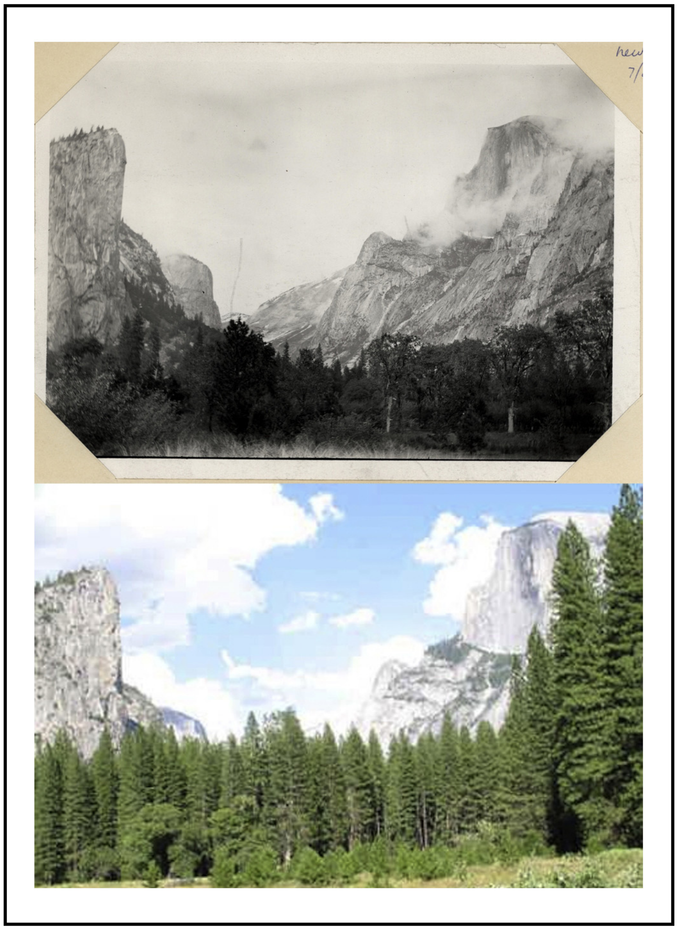

Figure 6. Taken almost a century apart, these two photos show a popular tourist destination within Yosemite National Park. Photographer: Joseph Grinnell. Date: 29 May 1911. MVZ photograph 486. Recent photograph, circa 2007, photographer: Grinnell Resurvey Team.

surveyed by the MVZ in the early twentieth century to document patterns of change in the fauna and flora. Detecting and analyzing this change involves documenting if and how species have moved around the landscape.[45] In addition to studying museum specimens, the research uses a variety of statistical approaches to model the occurrences of species as they were described in the early field notes. This occupancy modeling approach was designed to enable unbiased comparisons between modern and historical data.[46] Photograph retakes are carefully staged to allow an analysis of how vegetation, for example, has changed over time (Figure 6). This approach has proven extremely fruitful and reinvigorated interest in the collections. It would be very difficult, however to conduct a similar kind of resurvey project in the year 2060, or 2070, because it would be more challenging to obtain photographic and field note records of the conditions in the 1960s or 1970s. Interestingly, the GRP also raises a number of important questions about the analytical problems that arise when working with multiple computer databases and programs, which often contain competing and/or conflicting assumptions about the world.[47]

Now that the technologies are available, researchers who conducted field work during this period are thinking through how to prepare their photographs and notes for archival purposes. Recently, the MVZ began scanning field notes and established curatorial procedures to capture and archive any digital image, along with its accompanying metadata, and to make all of the data (image and metadata) available through their integrated collection database (Arctos). One of the biggest hurdles to making this happen is the deluge of data, a natural history problem with deep historical roots.[48] Patton estimates, for example, that there are more digital photographs waiting to be curated from the seven-year GRP as there were generated in the 31-year tenure of Grinnell as MVZ director.[49]

The early digitization of the MVZ's collections raises important questions about both the short- and long-term results of generating computer databases. To better understand the long-term results it would be necessary to examine Taxir's migration to a relational database in the 1990s and finally to an online, open-access, multi-institutional relational database in the 2000s – a period of rapid technological change that warrants further attention. The MVZ's database continues to expand its reach and connections with multiple institutions through the development of VertNet (vertnet.org), a massive international effort that is bringing together many vertebrate databases, in conjunction with new applications, with the aim of transforming how vertebrate biodiversity data can be used. Scanning the VertNet site, it is evident that computer databases reconfigured research over the long-term, but looking at the how these databases were first introduced challenges the revolutionary narrative.[50] Electronic data processing did not fundamentally change research overnight. In the short-term, digitization subtly affected workflow by enabling the efficient extraction of mass data, creating new ways to access data, and establishing new procedures for cataloging data. It also spurred a shift in the perceived scientific value of different kinds of objects and data. This shift deemphasized the importance of field notes, photographs, and in the 1990s, even the tangible specimens began to lose value as people began to interact with the specimen data online rather than physically examine specimens.

A deeply important short-term impact of computerization is best described as symbolic. To the larger scientific community, electronic data processing was seen as the way of the future and, in a sense, computerizing collections helped to insure the MVZ's future. A computerized database had the ability to mark an institution as innovative and research-oriented. Early computerization efforts met

[45] Craig M. Moritz, James L. Patton, Chris J. Conroy, Juan L. Parra, Gary C. White and Steven Beissinger. 2008. Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA. *Science* 322: 261–264.

[46] Morgan W. Tingley and Steven R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives an old data. *Trends in Ecology and Evolution* 24: 625–633.

[47] Ayelet Shavit and James R. Griesemer. 2009. There and Back Again, or the Problem of Locality in Biodiversity Surveys, *Philosophy of Science* 76: 273–294; Shavit and Griesemer. 2011. 'Transforming Objects into Data: How Minute Technicalities of Recording' Species Location 'Entrench a Basic Challenge for Biodiversity,' in *Science in the Context of Application* eds. Martin Carrier and Alfred Nordmann. *Boston Studies in the Philosophy of Science* 274:169–193.

[48] Staffan Müller-Wille and Isabelle Charmantier. 2012. Natural history and information overload. *Studies in the History and Philosophy of Biological and Biomedical Sciences* 43: 4–15; Strasser, 2012.

[49] Patton to Sunderland, 19 September 2012.

[50] Agar, 2006.

the MVZ's immediate need to modernize. More broadly, the ASM's efforts to build a larger digital network are emblematic of how the museum community began to think more collaboratively about long-term goals. Although networks have long been at the core of collections-based research, computerizing collections opened the possibility of creating new kinds of data networks and helped to frame the common, long-term goal of making natural history collections into a shared national resource.

Acknowledgements

I am grateful to Elihu Gerson, James L. Patton, Bruno J. Strasser and the Endeavor reviewers for helpful comments on earlier drafts and to the librarians at the Bancroft Library and the Smithsonian Institution Archives who provided assistance with the archives. Thank you to Chris Conroy for guiding me through the curatorial procedures for the MVZ's mammal collection. Thanks especially to Jim Patton for patiently and candidly sharing his many experiences. This material was based upon work supported by the National Science Foundation under Grant No. 0823401.