Collecting Nature: Practices, Styles, and Narratives

Strasser, Bruno J.

Collecting Nature:
Practices, Styles, and Narratives

by Bruno J. Strasser[*]

ABSTRACT

The standard narrative in the history of the life sciences focuses on the rise of experimentalism since the late nineteenth century and the concomitant decline of natural history. Here, I propose to reexamine this story by concentrating on a specific set of material and cognitive practices centered on collections. I show that these have been central for the production of knowledge not only in natural history, from the Renaissance to the present, but also in the experimental sciences. Reframing the history of the life sciences in this way makes historical continuities visible and raises new possibilities to contextualize recent developments in science, such as the proliferation of databases and their growing use.

INTRODUCTION

The "data deluge" (sometimes the "data tsunami") is upon us, and the resulting "data storms" and "floods" threaten to "drown" all those who have not learned to "swim in a sea of data."[1] From the pages of the magazine *Wired* to those of the *Economist*, *Nature,* and *Science,* an abundance of aquatic (and biblical) metaphors describe a new era in which humanity seems to be threatened by an unprecedented amount of data. The July 2008 cover of *Wired* went so far as to announce that the so-called deluge signified "the end of science" and to explain that "the quest for knowledge used to begin with grand theories," but now "it begins with massive amounts of data."[2] Numerous scientists in the natural and social sciences have announced the coming of age of a

* Yale University, Program in the History of Science and Medicine. Current address: University of Geneva, IUFE, Pavillon Mail, 40, Bd du Pont-d'Arve, CH—1211 Geneva 4; bruno.strasser @unige.ch.

I would like to thank Robert E. Kohler and Kathryn Olesko for their encouragement, comments, and infinite patience, and Robin Scheffler, Rachel Rothschild, and especially Helen Curry, as well as two anonymous referees for *Osiris*, for helpful suggestions.

[1] The expression "data deluge" is widely used, at least since the early 1990s; see P. Aldhous, "Managing the Genome Data Deluge," *Science* 262 (1993): 502–3, and more recently, G. Bell, T. Hey, and A. Szalay, "Computer Science: Beyond the Data Deluge," *Science* 323 (2009): 1297–8. A special report featured on the cover of the *Economist* was titled "The Data Deluge and How to Handle It" (February 27, 2010); *Wired* titled its issue on the subject "The End of Science" (July 16, 2008); *Nature* had a cover on "Big Data" (vol. 455, no. 7209 [2008]). For "data tsunami," see Anthony J. G. Hey, Stuart Tansley, and Kristin Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, Wash., 2009), 117, 131; for "floods," Bell, Hey, and Szalay, "Computer Science," 1297; for "swim in a sea of data," David S. Roos, "Bioinformatics—Trying to Swim in a Sea of Data," *Science* 291 (2001): 1260–1, on 1260.

[2] *Wired*, "End of Science" (cit. n. 1), cover.

"data-driven science" representing a "fourth paradigm" that follows the empirical, theoretical, and computational paradigms.[3] The value of data-driven science is being vigorously debated and contested in the scientific community. These discussions are all the more relevant because science funding agencies, at least in the United States, rely explicitly on a "hypothesis-driven" model of scientific research in evaluating research grant proposals.[4] The key assumption shared by almost all participants in the current debate over the value of data-driven sciences is that experimentation is the benchmark for what constitutes science. As the director of a National Institutes of Health (NIH) biomedical center put it in *Science*, the hypothesis-driven experimental method is "the scientific method," which "has driven conceptual inquiry for centuries and still forms the basis of scientific investigation."[5]

This assumption is also embedded in a narrative that dominates "generalist" histories of the life sciences.[6] According to this standard story, the study of living nature, from ancient Greece to the late Enlightenment, focused on naming, classifying, and describing the outer and inner morphology of plants and animals (*nommer, classer, décrire*, as Georges Cuvier put it). This project, known as "natural history" since at least the ancient Roman days of Pliny the Elder, initially was not concerned with change over time, as the modern sense of "history" would imply, but with the establishment of a "systematic" (and static) account of nature.[7] This project developed most forcefully during the Renaissance—thanks to the expansion of travel and the development of print—and found its home in wonder cabinets, gardens, herbaria, and eventually in natural history museums. In the mid-nineteenth century, this project was challenged by the rise of experimentalism, especially in physiology and embryology, which proposed to reveal the functions of life in the laboratory through the use of instruments and controlled experiments. Experimentation had occasionally played a role in the production of biological knowledge prior to that time—for example, in the seventeenth century in the work of William Harvey and Santorio Santorio, in the eighteenth with Abraham Trembley and Charles Bonnet—but unlike in the physical sciences, it remained marginal until the nineteenth century.[8] Then experimentalism began to displace natural history as the center of the life sciences. After 1900, the development of genetics, biochemistry, and crystallography hailed the newly recognized power of experimentation in unlocking the secrets of life and led to the ultimate triumph of experimentalism in the life sciences with the rise of molecular biology and the demise of natural history.

This narrative captures crucial aspects in the history of the life sciences since the late nineteenth century, such as the growing authority of the experimentalist discourse and the rise of particular experimental sciences. Yet, like its sister narrative

[3] For the social sciences, see Gary King, "Ensuring the Data-Rich Future of the Social Sciences," *Science* 331 (2011): 719–21.

[4] M. A. O'Malley et al., "Philosophies of Funding," *Cell* 138 (2009): 611–5.

[5] G. J. Nabel, "The Coordinates of Truth," *Science* 326 (2009): 53–4, on 53.

[6] E.g., Lois N. Magner, *A History of the Life Sciences* (New York, 2002); Peter J. Bowler and Iwan Rhys Morus, *Making Modern Science: A Historical Survey* (Chicago, 2005); Peter Dear, *The Intelligibility of Nature: How Science Makes Sense of the World* (Chicago, 2006).

[7] On the history of natural history, see Nicholas Jardine, James A. Secord, and Emma C. Spary, eds., *Cultures of Natural History* (New York, 1996); Paul Lawrence Farber, *Finding Order in Nature: The Naturalist Tradition from Linnaeus to E. O. Wilson* (Baltimore, 2000).

[8] Mirko Grmek, *La première révolution biologique: Réflexions sur la physiologie et la médecine du XVIIe siècle* (Paris, 1990).

on the laboratory revolution in medicine, it invites renewed questioning. Indeed, like all big narratives, this one excludes or obscures particular historical events whose significance might look different today than it did almost half a century ago when this narrative was first proposed. Detailed historical studies on the development of recent biological science (e.g., the changes that have led to the recent data deluge) offer good opportunities to test the explanatory power of this narrative, which was largely crafted from historical material about the specific changes that took place at the juncture of the nineteenth and twentieth centuries. Further motivating a reassessment of this narrative is the fact that some of the historical assumptions (the progress of experimentalism) and historiographic categories (experimentalism, morphology, and natural history) embedded in the narrative seem less obvious today than they did a few decades ago.

If historians of science want to remain relevant to current conversations about the changing nature of scientific research (including the data deluge and the data-driven sciences), they need to offer intellectual frameworks that can bring these changes into historical perspective, raise fresh questions, and challenge received views.[9] This article proposes such a framework, centered on the roles of collecting and collections in the production of knowledge in the life sciences. As I argue here, the development of the life sciences since the second half of the twentieth century is at odds with the standard narrative of the rise of experimentalism. Too much recent biological research, centered on the collection, comparison, and computation of biological data, does not fit well in a story focused on the triumph of experimentation over other ways of producing knowledge. To make sense of these recent historical transformations, I propose a different set of analytical categories, focused on a specific cognitive and material practice centered on the constitution and use of collections, which can in turn help us revisit the longer history of the life sciences, from the Renaissance to the present. It can also, as I suggest in the conclusion, offer promising venues to draw new connections between the life sciences and other sciences in which collecting and collections have played an equally important role. This offers new vistas to revisit the history of the sciences more generally, and in particular to view more critically the claims that data-driven science represents a turning point in the history of science.

GENERALIST VISIONS IN THE LIFE SCIENCES

Writing a generalist or "big-picture" history of biology poses particular challenges.[10] To begin with, despite the many books on the subject, there was no such word as *biology* for most of the subject's history. The study of nature was pursued from many different avenues, sometimes grouped under the name *natural history*, but the extension and meaning of this term has changed profoundly since the Renaissance. The word *biology* began to be used consistently around 1800, but did not enjoy wide currency until half a century later.[11] Institutionally, biology became a recognizable

[9] On the necessity to offer larger intellectual frameworks in the history of science in order to reach a broader readership, see Steven Shapin, "Hyperprofessionalism and the Crisis of Readership in the History of Science," *Isis* 96 (2005): 238–43.

[10] Ludmilla Jordanova, "Gender and the Historiography of Science," *Brit. J. Hist. Sci.* 26 (1993): 469–83; James A. Secord, "Introduction," *Brit. J. Hist. Sci.* 26 (1993): 387–9; Robert E. Kohler, "A Generalist's Vision," *Isis* 96 (2005): 224–9.

[11] Peter McLaughlin, "Naming Biology," *J. Hist. Biol.* 35 (2002): 1–4.

discipline in the United States from the late nineteenth century, but botanists and zoologists (and anthropologists and bacteriologists) in Europe usually enjoyed separate institutional and intellectual lives well into the mid-twentieth century.[12] Even after a comfortable institutional home was finally created for biology in universities, noisy infighting made amply clear that "biology" did not refer to a happy and united family. Its members disagreed on most things, including intellectual agendas, methodological approaches, and even the question whether it was a pursuit suited to both sexes.

The great diversity of biological practices certainly matched the diversity found within biology's objects of study, but this brings little comfort to the historian who would like to write a broad story of its development. Understandably, most historians of biology have fallen back on more limited projects and let themselves be guided by the disciplinary divisions of biology, such as physiology, embryology, immunology, systematics, evolution, and molecular biology. Writing within or about such categories makes sense: they were recognized by the historical actors themselves and structured their intellectual goals, research practices, social communities, and institutional homes. Yet taking these as units of analysis has tended to isolate historical subjects from other scientific endeavors and has limited historical explanations of their development to factors internal to a given discipline.

In the 1970s, a few historians attempted to write more encompassing stories. For example, William Coleman and Garland Allen each covered an entire century (the nineteenth and the twentieth, respectively) in their histories of biology and incorporated different approaches to the study of life into their analyses.[13] They organized their narratives around intellectual concerns such as "form," "function," and "evolution" (for Coleman), or disciplines such as "embryology" and "genetics" (for Allen). Overall, they told similar stories. In the late nineteenth century, studies of function (through experimentation) began to take over studies of form (i.e., "morphology"), and in the twentieth century, the experimental method triumphed in the life sciences, culminating in the success of molecular biology. For Coleman, "In its name—experiment—was set in motion a campaign to revolutionize the goals and methods of biology." For Allen, "It was the twentieth century that saw the fanning out of the experimental method in all areas of biology."[14]Although they differed in their explanation of what drove this transformation—the import of new chemical and physical methods from the "outside" (Coleman) or a revolt against morphology from the "inside" (Allen)—both focused on the broad category of experimentalism, as a material practice, a physical place, and an intellectual program. Coleman and Allen's story has served as a framework for most subsequent histories of the life sciences.

To fully understand the origins of the historical categories that structured the narratives proposed by Coleman and Allen, it is helpful to bring them into the context of the development of the life sciences at the time of their production. In the late 1960s and 1970s, when Coleman's and Allen's works were written, biology departments had been carved into battlegrounds between proponents of different research agendas, yet it was already clear that the future belonged to the experimental life sciences.

[12] William V. Consolazio, "Dilemma of Academic Biology in Europe," *Science* 133 (1961): 1892–6.

[13] Coleman, *Biology in the Nineteenth Century: Problems of Form, Function and Transformation* (Cambridge, 1971); Allen, *Life Science in the Twentieth Century* (Cambridge, 1978); Allen, "Morphology and Twentieth-Century Biology: A Response," *J. Hist. Biol.* 14 (1981): 159–76, as well as the other contributions to that special issue of the *Journal of the History of Biology*.

[14] Coleman, *Biology*, 2; Allen, *Life Science*, xvi (Both cit. n. 13).

Beginning in 1962, the Nobel Prize committee rewarded molecular biologists year after year for their experimental work. Universities around the world created institutes of molecular biology, testifying to the achievements and promises of this new science.[15] Molecular biologists launched journals devoted to their field, beginning with the *Journal of Molecular Biology* in 1959. Naturalists complained bitterly about the excessive (in their view) attention and support given to experimental approaches.[16] In his autobiography, revealingly titled *Naturalist,* the evolutionary biologist E. O. Wilson recalled the epic wars in Harvard's corridors and faculty meetings that pitted him against another young faculty member, the molecular biologist James D. Watson (the "most unpleasant person" Wilson had ever met).[17] Other naturalists, such as Ernst Mayr and George Gaylord Simpson, waged a war along the same battle lines (Allen was a student of Simpson at Harvard in 1966).[18] They criticized the new molecular evolutionists, who claimed to reconstruct the history of life through the comparison of single molecules, for being overly simplistic. In response, the molecular evolutionists ridiculed the traditional evolutionists, whose work focused on examination of skeletons and fossils, as outdated museum workers who toiled on characters too subjective to be of any scientific value. To mock their opponents' morphological methods and illustrate their limited scope, two molecular evolutionists asked in 1967, "How many vertebrae does a sponge have?"[19] Mayr attempted to draw up a peace settlement, which divided biology's territory into complementary fields, organismic and reductionist biology, concerned with asking historical and functional questions and answering them by appealing to ultimate and proximal causes, respectively.[20] The institutional and intellectual successes of molecular biology in the 1960s provided an end point to Coleman's and Allen's narratives, while the boundaries that were drawn by biologists themselves provided the outline for categories such as *experimentalist* and *naturalist*.

This is not to say that the methodological and disciplinary distinctions drawn by Coleman and Allen, based on their later twentieth-century experiences, were fully at odds with the distinctions made by scientists using the same terms at the turn of the century, nor that these distinctions did not capture some of the important tensions running though the life sciences around 1900. But seeing today what their categories meant to the state of the life sciences and the debates among life scientists in the 1960s invites us to consider a different reading of the broad history of the life sciences. This article is an experiment in *not* taking experimentalism as the end point

[15] On molecular biology as a postwar science, see Soraya de Chadarevian, *Designs for Life: Molecular Biology after World War II* (Cambridge, 2002); Bruno J. Strasser, "Institutionalizing Molecular Biology in Post-war Europe: A Comparative Study," *Stud. Hist. Phil. Biol. Biomed. Sci.* 33 (2002): 533–64.

[16] Michael R. Dietrich, "Paradox and Persuasion: Negotiating the Place of Molecular Evolution within Evolutionary Biology," *J. Hist. Biol.* 31 (1998): 85–111; Joel B. Hagen, "Naturalists, Molecular Biology, and the Challenge of Molecular Evolution," *J. Hist. Biol.* 32 (1999): 321–41.

[17] Wilson, *Naturalist* (Washington, D.C., 1994).

[18] Mayr, "Cause and Effect in Biology," *Science* 134 (1961): 1501–6; Mayr, "The New versus the Classical in Science," *Science* 141 (1963): 765; Simpson, "Organisms and Molecules in Evolution," *Science* 146 (1964): 1535–8; Theodosius Dobzhansky, "Biology, Molecular and Organismic," *American Zoologist* 4 (1964): 443–52; Dobzhansky, "Taxonomy, Molecular Biology, and the Peck Order," *Evolution* 15 (1961): 263–4.

[19] W. M. Fitch and E. Margoliash, "Construction of Phylogenetic Trees: 2. How Well Do They Reflect Past History?" *Brookhaven Symposia in Biology* 1 (1969): 217–42, on 238.

[20] Erika Lorraine Milam, "The Equally Wonderful Field: Ernst Mayr and Organismic Biology," *Hist. Stud. Nat. Sci.* (2010): 279–317.

of the story and *not* taking the contrasts drawn by Mayr between organisms and molecules, and by Allen and Coleman between morphology (and natural history) and experimentalism, as the defining tensions in the history of the life sciences.

GENERALIST VISIONS, MORE GENERALLY

Taking this approach to a generalist vision for the history of the life sciences creates new challenges. Unless one's goal is a history of great men, from Aristotle to, say, J. Craig Venter, it is necessary to choose some high-level analytical category to structure a narrative that will highlight the profound changes and continuities in the history of the life sciences. The goal is not to find categories that enable us, as Plato would suggest, to better carve history "at its joints" (the past is not a chicken), but to find categories that make meaningful connections between historical practices, actors, and events and that can become, at least temporarily, powerful heuristic tools to understand the past and the present. The most common categories used by historians of science have included periods (e.g., romanticism), spaces (the Atlantic world), nations (Germany), disciplines (physiology), ideas (extinction), styles (holistic), and places (laboratories). A combination of styles and places could be particularly promising to structure a generalist vision.

Styles of scientific reasoning, under various names, have been used productively by authors as different as Ludwig Fleck (*Denkstil*), Thomas Kuhn ("paradigms"), Michel Foucault ("episteme"), Gerald Holton ("themata"), Alistair Crombie ("styles"), Jon Harwood ("national styles"), John Pickstone ("ways of knowing"), and others.[21] These analytical categories help historians make sense of the development of science, capturing some of its methodological texture with more nuance than a unique and atemporal "scientific method" would allow. Crombie, for example, distinguishes six styles of scientific reasoning, including elementary postulation, comparative ordering, analogical modeling, statistical analysis, historical derivation, and experimental control.[22]

Because they focus on cognitive practices in science, narratives structured around styles usually pay little attention to material practices and thus tend to separate activities that are linked by common material practices (e.g., agricultural breeding and academic genetics) if they reflect different cognitive goals. Some authors, such as Pickstone, have considered this difficulty and have attempted to use categories that reflect to some extent both the cognitive and the material.[23] "Analysis," one of Pickstone's ways of knowing, for example, is understood as both a mental operation of analyzing abstract ideas into more specific components and a material operation of physically dividing scientific objects into their constitutive parts. Another strength of Pickstone's approach is that even though his ways of knowing each have their own historicity and have, for example, enjoyed their greatest successes at different times,

[21] For a discussion of Pickstone's categories, see Bruno J. Strasser and Soraya de Chadarevian, "The Comparative and the Exemplary: Revisiting the Early History of Molecular Biology," *Hist. Sci.* 49 (2011): 317–36.

[22] Crombie, *Styles of Scientific Thinking in the European Tradition: The History of Argument and Explanation Especially in the Mathematical and Biomedical Sciences and Arts*, 3 vols. (London, 1994).

[23] Pickstone, "Museological Science? The Place of the Analytical/Comparative in Nineteenth-Century Science, Technology and Medicine," *Hist. Sci.* 32 (1994): 111–38; Pickstone, *Ways of Knowing: A New History of Science, Technology and Medicine* (Manchester, 2000); Pickstone, "Working Knowledges before and after circa 1800: Practices and Disciplines in the History of Science, Technology and Medicine," *Isis* 98 (2007): 489–516.

they do not replace each other, as do Kuhnian paradigms, but add new layers to the makeup of science, technology, and medicine. An approach like Pickstone's, which considers styles of reasoning as individual components, makes it possible to analyze scientific practices in terms of these different elements, present in various proportions, without reducing scientific practices to any one of them, like Crombie's styles do. Chunglin Kwa's recent history of science uses Crombie's taxonomy to *classify* different sciences into his six styles; by contrast, I propose here, in line with Pickstone's ways of knowing, to *analyze* different sciences in terms of the material and cognitive practices on which they rely, focusing specifically on practices centered on collections.[24] This approach illuminates the heterogeneity of cognitive and material practices within disciplines and the similarities among disciplines, whereas disciplinary histories have tended to stress the unity of cognitive and material practices within disciplines and the differences among disciplines. One does not need to adopt Pickstone's categorization of the various ways of knowing to appreciate the value of analyzing sciences according to different kinds of practices, such as "collecting" and "comparing."

This is not necessarily a simple task: identifying the role of epistemic practices from the testimonies of historical actors can be challenging. In this respect, life-science historians have faced the same problem as medical historians exploring the "laboratory revolution" in medicine.[25] A standard account, supported by abundant documentary evidence, gave voice to the physicians who, since the late nineteenth century, emphasized how much their profession had been transformed by the laboratory sciences. In short, they claimed that medicine had finally become "scientific," thanks to the introduction of experimental methods. But as a number of studies since the 1980s have convincingly shown, much of medical practice remained unaffected by the new experimental sciences.[26] This does not mean that the laboratory was irrelevant for medicine's transformation, simply that it was more of a rhetorical tool, at least in the United States, used by physicians for the social elevation of their profession. Similarly and at the same time, the repeated claims by scientists about the superiority of the experimental over the natural historical method did not necessarily reflect actual changes in their research practices. To be sure, Coleman, Allen, and many others have amply documented that the discourse about the power of experimentation was not mere rhetoric. It did indeed capture a significant transformation and reflected the growing importance of those sciences that relied most on experimental methods, such as genetics or biochemistry. Nevertheless, a key question remains: Did the unquestionable intellectual successes of the experimental sciences in the twentieth century result from their reliance on the experimental approach alone and its superiority over other forms of inquiry, such as those based on collections? It is difficult to assess solely from the statements made by scientists that this approach had played some role in the successes of their investigative enterprises, because in a cultural context where the association with experimentation carried such authority and prestige and where other approaches, such as those based on collections, were derided as old-fashioned,

[24] Kwa, *Styles of Knowing: A New History of Science from Ancient Times to the Present* (Pittsburgh, 2011).

[25] Andrew Cunningham and Perry Williams, *The Laboratory Revolution in Medicine* (Cambridge, 1992).

[26] John Harley Warner, "Ideals of Science and Their Discontents in Late Nineteenth-Century American Medicine," *Isis* 82 (1991): 454–78.

one can easily understand how researchers might have downplayed the role of the latter in their work, or even denied that it played any significant role at all.

To explore the possibility that collecting practices have been important, not only for the field and museum sciences but also for the laboratory sciences, it is first necessary to question the common conflation by historical actors and scholars alike of places and practices. The "geographical turn" in science studies brought a welcome focus on neglected issues, such as the circulation of people, things, and knowledge,[27] but unfortunately, it has also reified places, such as laboratories and museums, as spaces where similar kinds of practices are supposed to have occurred.[28] It is too often assumed that the laboratory was necessarily a place of experimentation, in the same way that the field was a place of collection and the museum a place of comparison. Even though these have been the key practices in each of these three places, all witnessed a variety of other practices carried out in service of scientific research. The laboratory, in the nineteenth century alone, was a place for teaching morphology, for preparing specimens, and for conducting experiments;[29] the field was a place of collecting, but also of experimentation, observation, and other practices;[30] and the museum (the key target of ridicule for the proponents of the laboratory) hosted a variety of epistemic practices, including experimentation. For example, in 1928, the American Museum of Natural History established its own laboratory devoted to experimental research.[31] Surprisingly, historical studies that take "place" as their focus have not fully taken advantage of one of cultural geography's key insights: places are not passive receptacles for practices; instead, practices generate places (i.e., a café is a place where people drink coffee). This insight offers promising venues to develop analytical categories connecting places and practices, paying attention to both.[32]

A NEW GENERALIST VISION FOR THE LIFE SCIENCES?

The remainder of this article focuses on a specific set of epistemic and material practices, based on collecting things and data, for the constitution of collections, and their use, especially as tools for comparative studies. This account takes a broad view from the Renaissance to the present, but emphasizes the developments in the twentieth century. Few would challenge the fact that from the fifteenth to the nineteenth centuries,

[27] David N. Livingstone, *Putting Science in Its Place: Geographies of Scientific Knowledge* (Chicago, 2003); D. A. Finnegan, "The Spatial Turn: Geographical Approaches in the History of Science," *J. Hist. Biol.* 41 (2008): 369–88; Simon Naylor, "Introduction: Historical Geographies of Science—Places, Contexts, Cartographies," *Brit. J. Hist. Sci.* 38 (2005): 1–12; S. Shapin, "Placing the View from Nowhere: Historical and Sociological Problems in the Location of Science," *Trans. Inst. Brit. Geogr.* 23 (1998): 5–12.

[28] Ian F. McNeely and Lisa Wolverton, *Reinventing Knowledge: From Alexandria to the Internet* (New York, 2008).

[29] On the different uses of the laboratory, see the "Focus" section "Laboratory History," *Isis* 99 (2008): 761–802.

[30] Robert E. Kohler, *Landscapes and Labscapes: Exploring the Lab-Field Border in Biology* (Chicago, 2002).

[31] Gregg Mitman and Richard W. Burkhardt Jr., "Struggling for Identity: The Study of Animal Behavior in America, 1930–1950," in *The Expansion of American Biology*, eds. Keith R. Benson, Jane Maienschein, and Ronald Rainger (New Brunswick, N.J., 1991), 164–94.

[32] On places and practices, see Henri Lefebvre, *The Production of Space* (Oxford, 1991). For a successful attempt at this in the history of science, see Robert E. Kohler, "Place and Practice in Field Biology," *Hist. Sci.* 40 (2002): 189–210, and Richard W. Burkhardt Jr., "The Leopard in the Garden: Life in Close Quarters at the Muséum d'Histoire Naturelle," *Isis* 98 (2007): 675–94.

collections such as cabinets of curiosities and wonder cabinets, museums and zoos, and gardens and herbaria were all, in widely different ways, central to the production of knowledge about nature. The key historical question is, What happened to collecting practices after these institutions became less prominent in academic science? The standard account emphasizes that as the experimental sciences (such as genetics, biochemistry, and eventually molecular biology) gained importance, collecting practices and collections associated with museums and other institutions devoted to the pursuit of natural history lost their central role in the production of knowledge in the life sciences. This article argues that, on the contrary, collecting and collections have remained essential in the experimental sciences for the production of biological knowledge. The proliferation of databases of experimental data is the most visible sign of the key role played by collections in the life sciences (and elsewhere). If this is so, one can explore the development of the life sciences from the early modern period to the present in a new way.

One might immediately object that early collections of specimens differ too much from twenty-first-century "biobanks" (such as collections of molecules) or databanks (such as collections of data about molecules) to be productively compared to them. One should remember, however, that early modern and modern collections were not simply storehouses for whole organisms. Zoological museums often stored only bones and skins, and herbaria only parts of plants. Other collections of plant and animal parts became particularly important beginning in the nineteenth century, especially collections of seeds, blood, tissues, and cells.[33] Furthermore, natural history collections always included data, in the form of drawings or verbal descriptions of species, alongside material specimens. As Martin Rudwick has so beautifully shown, Cuvier's work relied not only on the fossils and bones present at the Muséum d'Histoire Naturelle in Paris, but also on the drawings of his own "paper museum."[34] Thus, it should not seem too big a step to argue that collections of data about biological molecules (such as DNA sequences, protein-structure coordinates, or functional MRI images), when gathered in electronic databases, can be subsumed under the same category of *collection* as collections of plants, animals, fossils, and so on. This is an expansion of the usual meaning of *collection* in the life sciences, but it is one that seems not only conceptually plausible, but also historiographically useful.

A more detailed historical, epistemological, and ontological justification for subsuming all of these—museums, herbaria, biobanks, and databases—under a unique category lies beyond the scope of this article. For the time being, let us assume that these assemblages share a sufficient number of similarities to be plausibly brought together under the single analytical category of *collection*, and that this, in conjunction with practices of collecting and comparing, can be used as the basis to reexamine the history of the life sciences. As I will show, foregrounding the history of collections, collecting, and comparing, instead of the categories of museum and laboratory, or natural history and experimentation, offers a different perspective on the history of the life sciences. It allows us to make connections and see continuities otherwise obscured, to ask new questions about the transformations of the life sciences in the

[33] Jack Ralph Kloppenburg, *First the Seed: The Political Economy of Plant Biotechnology, 1492–2000* (Cambridge, 1988); Susan E. Lederer, *Flesh and Blood: Organ Transplantation and Blood Transfusion in Twentieth-Century America* (Oxford, 2008); Bronwyn Parry, *Trading the Genome: Investigating the Commodification of Bio-information* (New York, 2004).

[34] Rudwick, "George Cuvier's Paper Museum of Fossil Bones," *Arch. Natur. Hist.* 27 (2000): 51–68.

twentieth century, and to use the rich historical literature on collections up to the nineteenth century to explore and inform our understanding of recent developments.

THE PRACTICE OF COLLECTING

To show the connections between the practice of collecting as it occurred in the twentieth century and in earlier centuries, I will now provide a cursory and necessarily incomplete overview of collecting in natural history from the Renaissance to around 1900 (later periods are territories largely uncharted by historians). Constrained by space and the limited existing literature, I address only three questions: How was it done? By whom? And for what purpose?

One of the first volumes to focus on collecting as a way of knowing, *Sammeln als Wissen,* brought several contributors together to reflect on particular collecting practices.[35] All of them were associated with early modern natural history, yet most of the contributions fell back on the traditional themes of museum history—the exploration of the collections themselves and their cultural meanings—rather than considering collecting as a practice. The richest volume on collecting practices in the life sciences, *Cultures of Natural History*, offers unique insights to understand natural history collecting, but only until the nineteenth century.[36] Some equally revealing examples of particular collecting practices from the mid-nineteenth to the twentieth century include studies of naturalists, especially botanists, in Britain; the Smithsonian's scramble to secure artifacts from American Indians in the Northwest; geologists' search for fossils in England; the efforts of museums and government agencies to survey the biodiversity of the United States; and the attempts of medical researchers to obtain brains from kuru patients in Papua New Guinea.[37] Each study highlights different aspects of collecting, but taken together, they can serve as a basis for outlining some of the common features of collecting.

Robert Kohler seems to be the first historian to have explicitly attempted to analyze collecting as a practice and to focus on the modern, rather than the early modern, period.[38] He argues that "collecting sciences" have included not only systematic biology, but also anthropology and ethnology, geology and mineralogy, and even, at some point, pathology and chemistry. He resolutely takes "a generous view of collecting practices" and convincingly argues that the collecting sciences should not be limited to the natural sciences, but should also include the social sciences and other sciences involved in collecting data.[39] Nonetheless, he draws a sharp distinction between the collecting sciences that focus on things and those that do not. Kohler sees the most characteristic aspect of collecting sciences as being the materiality, the

[35] Anke te Heesen and Emma C. Spary, eds., *Sammeln als Wissen: Das Sammeln und seine Wissenschaftsgeschichtliche Bedeutung* (Göttingen, 2001).

[36] Jardine, Secord, and Spary, *Cultures of Natural History* (cit. n. 7).

[37] David Elliston Allen, *The Naturalist in Britain: A Social History* (London, 1976); Douglas Cole, *Captured Heritage: The Scramble for Northwest Coast Artifacts* (Norman, Okla., 1995); Simon J. Knell, *The Culture of English Geology, 1815–1851: A Science Revealed through Its Collecting* (Aldershot, 2000); Robert E. Kohler, *All Creatures: Naturalists, Collectors, and Biodiversity, 1850–1950* (Princeton, N.J., 2006); Warwick Anderson, *The Collectors of Lost Souls: Turning Kuru Scientists into Whitemen* (Baltimore, 2008).

[38] Kohler, "Finders, Keepers: Collecting Sciences and Collecting Practice," *Hist. Sci.* 45 (2007): 428–54.

[39] Kohler, *All Creatures* (cit. n. 37), 433.

"thing-y" nature, of the objects it deals with. He defines collecting scientists as not just "finders," because "all scientists are finders (in one way or the other)," but as "keepers," because "only collecting scientists are also keepers."[40] For Kohler, finding material objects and keeping them defines the collecting sciences. Although I take issue with some aspects of Kohler's definition, especially with respect to the "found" nature of things, my attempt to characterize collecting practices takes a similarly broad approach, from the fifteenth to the end of the nineteenth century.

Naturalists throughout this period are generally recognized as having been active collectors. Although their collecting practices took extremely diverse forms, the key challenge of collecting, and of establishing a collection, remained the same over time: how to bring spatially dispersed objects to a central location and make them commensurable. Consequently, collecting was (and is), above all, a spatial practice. Renaissance collections were filled first with objects coming from the immediate environment.[41] Local plants, animals, and minerals, especially, were brought into closer proximity with one another. Collectors, after having exhausted the diversity of their local surroundings, embarked on the more ambitious goal of filling their collections with objects found far beyond their everyday reach.

Establishing this kind of collection, like establishing empires, required the mastery of space. Collectors produced a movement of natural things, which were often dispersed across the world, toward central locations, just as empires produced movements of goods from colonies to metropoles. Unsurprisingly, colonial powers were collecting powers, and colonies constituted rich collecting grounds.[42] The geographical reach of an empire represented an immense field for collecting. The objects in the collections of Kew Gardens or the British Museum, for example, came from the same places and followed the same routes as the other goods circulating through the British Empire. And collecting, just like the imperial enterprise, required domination over people, not just things. Indeed, most collecting was done by proxy. Collectors in the metropole relied on local naturalists, hunters, and gatherers in the colonies, although some collectors did go into the field themselves to collect specimens (and were sometimes carried on comfortable chairs by the locals).[43]

But collecting should not be reduced, as it sometimes has been, to the history of colonial exploitation. It also followed the lines of gift economies, as in Renaissance Italy or the French republic of letters.[44] For example, in eighteenth-century France,

[40] Ibid., 432.

[41] Paula Findlen, *Possessing Nature: Museums, Collecting, and Scientific Culture in Early Modern Italy* (Berkeley and Los Angeles, 1994); Brian W. Ogilvie, *The Science of Describing: Natural History in Renaissance Europe* (Chicago, 2006).

[42] Lucile H. Brockway, *Science and Colonial Expansion: The Role of the British Royal Botanic Gardens* (New York, 1979); Richard W. Burkhardt Jr., "Naturalists' Practices and Nature's Empire: Paris and the Platypus, 1815–1833," *Pacific Sci.* 55 (2001): 327–41; Londa L. Schiebinger and Claudia Swan, eds., *Colonial Botany: Science, Commerce, and Politics in the Early Modern World* (Philadelphia, 2005); Daniela Bleichmar and Peter C. Mancall, eds., *Collecting across Cultures: Material Exchanges in the Early Modern Atlantic World* (Philadelphia, 2011).

[43] Londa L. Schiebinger, *Plants and Empire: Colonial Bioprospecting in the Atlantic World* (Cambridge, Mass., 2004), chaps. 1–2. See also Harold John Cook, *Matters of Exchange: Commerce, Medicine, and Science in the Dutch Golden Age* (New Haven, Conn., 2007).

[44] Paula Findlen, "The Economy of Scientific Exchange in Early Modern Italy," in *Patronage and Institutions: Science, Technology, and Medicine at the European Court, 1500–1750*, ed. Bruce T. Moran (Rochester, N.Y., 1991), 5–24; Findlen, *Possessing Nature* (cit. n. 41).

where the possession of natural objects became a sign of distinction, there were coveted cultural objects to be offered to collectors, often rich patrons who would reciprocate by offering other natural objects or patronage in return.[45] These practices created what Emma Spary has so appropriately called, referring to the network of collectors around the French Jardin des Plantes in the eighteenth century, a system of "polite indebtedness."[46] Furthermore, although collecting centers have often been imperial capitals (e.g., London, Amsterdam, and Paris), this was not always the case (e.g., Geneva, Kew, and Montpellier). It might thus be more productive to think that collections only became "centers" once they succeeded in generating a "periphery." By convincing naturalists around the world to send them specimens, museums and botanical gardens became centers for the production of natural knowledge.

Natural history objects not only traveled as gifts along existing social networks, but also moved as commodities that could simply be purchased by collectors. This practice of collecting was particularly important for those who wished to obtain specimens beyond the frontiers of the empire or who did not have access to colonial networks of power. In the busy merchant port at Canton in the nineteenth century, for example, British collectors purchased animals and plants from the luxuriant markets, and so the city itself became the "field" for these second-order collectors.[47] At the same time in the United States, animal dealers, often hunter-entrepreneurs, offered wild animals for sale to zoos and natural history museums.[48]

When the routes of empire or commerce were unavailable, collectors mounted their own expeditions. In the late nineteenth and early twentieth centuries, the major natural history museums in Europe and in the United States commissioned expeditions to Asia, Africa, and South America. The American Museum of Natural History, for example, sent groups of naturalist-collectors to Congo between 1909 and 1915 to survey the local fauna and bring back mammals for the museum's African Hall.[49] Expeditions were not always successful in bringing back animals, especially live animals, but they always succeeded in returning with stories, and these stories shaped the image of the naturalist-collector into one of an explorer-adventurer.[50] In the same period, as Kohler has shown, a new way of collecting developed under the auspices of museums and governmental agencies: systematic survey collecting. Thanks to a specific nexus of environmental, technological, and cultural factors, this produced an "inner frontier," where biologically rich collecting spaces were never too far from civilization.[51]

Many collectors associated with these different modes of collecting were what Londa Schiebinger has aptly called "armchair collectors," in that they relied exclu-

[45] Philipp Blom, *To Have and to Hold: An Intimate History of Collectors and Collecting* (Woodstock, N.Y., 2003).

[46] Spary, *Utopia's Garden: French Natural History from Old Regime to Revolution* (Chicago, 2000), 77.

[47] Fa-ti Fan, *British Naturalists in Qing China: Science, Empire, and Cultural Encounter* (Cambridge, Mass., 2003).

[48] On the animal dealers, see Mark Barrow, "The Specimen Dealer: Entrepreneurial Natural History in America's Gilded Age," *J. Hist. Biol.* 33 (2000): 493–534; on hunters, Elizabeth Hanson, *Animal Attractions: Nature on Display in American Zoos* (Princeton, N.J., 2002), chap. 3.

[49] Lyle Rexer and Rachel Klein, *American Museum of Natural History: 125 Years of Expedition and Discovery* (New York, 1995).

[50] Douglas J. Preston, *Dinosaurs in the Attic: An Excursion into the American Museum of Natural History* (New York, 1986).

[51] Kohler, *All Creatures* (cit. n. 37).

sively on a network of individuals to gather materials for them.[52] Some had participated in collecting expeditions in the field early in their careers, but later became curators at institutions such as natural history museums or botanical gardens, where they directed their collecting enterprises from their desks through correspondence networks. These armchair collectors painstakingly attempted to coordinate and discipline field collectors in making standardized observations of the specimens they collected, a condition for successfully creating a reliable "collective observer" and a useful collection.[53] The fact that such collectors often published in their own name faunas and floras that were based on specimens and data gathered by numerous collaborators maintained the illusion that the knowledge produced through collecting practices was, like that produced through experimentation, the result of an individualist endeavor. This conformed to the persistent ideal that only individuals are the creators of knowledge.[54]

Because collecting was essentially a collective practice, carried out by very different actors, the issue of epistemic and social coordination was essential.[55] When Western physicians collected brains from patients who had recently died of kuru in New Guinea, they had to negotiate over the status of the brain and their own status as collectors; they were considered physicians by some, sorcerers by others.[56] Similarly, when envoys from the Smithsonian collected canoes and other artifacts on the Northwest coast, they bargained over the authenticity and value of these artifacts with the natives, who had sometimes created them especially for the collectors.[57] Even when all the individuals in the collective were of a similar professional background—botanists, for example, in the case of the Kew Gardens naturalist Joseph Hooker's collecting plants from his New Zealand correspondents—intense arguments took place between the field and the institutional collector over the status of a rare find and its relation to the attribution of credit and authorship.[58]

Given this diversity of modes of collecting, what were the main characteristics of collecting as a unified practice? First, collecting was a spatial practice, always negotiating problems of position, scale, and reach. Second, and following on the first, collecting was a local practice. Even global collecting efforts, at some level, required local collectors to gather materials in the field. Third, it was a collective practice. Few individuals assembled their collections alone; almost all relied on extended networks of people. As Spary has put it, natural history in the eighteenth century, a science highly reliant on collecting, was "a science of networks."[59] Finally, given the heterogeneity of these networks, bringing together naturalists, hunters, and merchants (to

[52] Schiebinger, *Plants and Empire* (cit. n. 43), chap. 1.

[53] Peter Galison and Lorraine Daston, "Scientific Coordination as Ethos and Epistemology," in *Instruments in Art and Science: On the Architectonics of Cultural Boundaries in the 17th Century*, ed. Helmar Schramm, Ludger Schwarte, and Jan Lazardzig (Berlin, 2008), 296–333.

[54] Mario Biagioli and Peter Galison, eds., *Scientific Authorship: Credit and Intellectual Property in Science* (New York, 2003).

[55] For a useful perspective on the translation of various interests in a collecting network, see Susan Leigh Star and James R. Griesemer, "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939," *Soc. Stud. Sci.* 19 (1989): 387–420.

[56] Anderson, *Collectors of Lost Souls* (cit. n. 37).

[57] Cole, *Captured Heritage* (cit. n. 37).

[58] Jim Endersby, *Imperial Nature: Joseph Hooker and the Practices of Victorian Science* (Chicago, 2008).

[59] Spary, *Utopia's Garden* (cit. n. 46), 97.

name a few) required the translation of very diverse interests. Because the objects of collections meant very different things to these different people, collectors negotiated complex issues of credit, translating among often incommensurable values to keep objects flowing toward their collection.

WHO THE COLLECTORS WERE

Still unanswered in this overview of collecting as a practice are the questions, Who was doing the collecting, how did these individuals characterize themselves, how were they characterized by others, and what was their social position among scientists? Answering these questions offers illuminating material for comparisons between old and new collectors. In his *Collectors and Curiosities*, Krzysztof Pomian provides a wonderfully rich account of the world of collectors—a group that included far more than natural history collectors—in France and Italy between the sixteenth and the eighteenth centuries.[60] He shows how collectors who had amassed medals, paintings, instruments, and natural specimens and stored these together in collections such as wonder cabinets began, in the seventeenth century, to specialize by focusing on a single kind of object. Antiquaries favored artifacts reflecting the life of the ancients; savants favored the collection of natural history objects. Within each category of collected object, the collectors were a highly heterogeneous mix of people with different interests, a characteristic that has continued to define collectors to the present day.[61] As a result, individual collectors have had very diverse and unstable social identities.

This diversity of social identities certainly applies to those who collected natural objects. Beginning in the Renaissance, when the social identity of "naturalist" was solidifying, most naturalists were collectors of some sort, but not all collectors of natural objects were naturalists. Natural objects were collected by all kinds of people for all kinds of reasons. But it is possible to identify some of the categories according to which collectors came to be identified. These included, since the nineteenth century at least, the amateur and the professional, and the field and the museum collector.

Since the nineteenth century, amateur naturalists, who often had as much expertise in their field of specialty as professionals, were indispensable to collecting enterprises, from surveys of local floras to expeditions to remote places where only local inhabitants possessed knowledge of their natural environment.[62] The enthusiastic participation of amateurs in natural history collecting proved to be a mixed blessing for the professional naturalists, such as Hooker at Kew Gardens. He could count on local collectors to provide specimens from the other side of the Earth (in this case, New Zealand), but tensions arose over issues of credit, especially over the right to name new species. Amateurs were generally unpaid, so they sought remuneration in other forms, such as the right to name species for posterity. But Hooker and other naturalists at the metropole thought that naming was the privilege of the

[60] Pomian, *Collectors and Curiosities: Paris and Venice, 1500–1800* (Cambridge, 1990).

[61] For a rich account of cabinet collecting in the Renaissance and early modern period, see Blom, *To Have and to Hold* (cit. n. 45).

[62] David Elliston Allen, "Amateurs and Professionals," in *The Cambridge History of Science,* vol. 6, *The Modern Biological and Earth Sciences*, ed. Peter J. Bowler and John V. Pickstone (Cambridge, 2009), 15–33.

professional who entered the specimen into the formal scientific literature.[63] The association between collecting and amateur science has also been at times a curse for naturalists in their quest for professional respectability in the sciences, especially where amateur collecting became an accepted leisure activity. For example, in Victorian Britain, collecting ferns was a popular hobby for the rising bourgeoisie, just as collecting plants became part of a middle-class ideal of vacationing in the United States at the turn of the twentieth century.[64] Because the late nineteenth century was also a moment when the sciences were becoming increasingly professionalized, the association of amateur activity with collecting was detrimental to the development of sciences dependent on collecting. As Kohler has put it, "Scientific collecting was just too much like camping and sport hunting to be taken seriously by guardians of the public purse—too much like plain fun."[65]

Collecting, especially in botany, was not only associated with amateurs; it was to some extent gendered as a female pursuit. Even before the Victorian days of the great fern craze, when women eagerly collected specimens, botanists tried to dispel the idea that in Britain the pursuit was "of so low a character, as to be calculated for the amusement of women," as one commentator put it in 1831.[66] Half a century later, the author of a letter published in *Science* was still trying to counter the widespread idea that botany and the collecting of botanical specimens were "suitable enough for young ladies and effeminate youths, but not adapted for able-bodied and vigorous-brained young men who wish to make the best use of their powers."[67] In the early twentieth century, when women became an increasingly important "workforce" in science, they were predominantly relegated to subaltern and repetitive tasks, such as data and specimen collection; this division of labor reinforced the gendering of the sciences that depended on collecting practices.[68] Into the twentieth century, the gendering of natural history collecting continued to affect all of biology. The molecular biologist Sydney Brenner, before the rise of his new discipline, stated, "Biology, I am sorry to say, was a subject for girls."[69]

The association of natural history collecting with amateurs limited the professional opportunities of collectors. Among professional collectors, a few found coveted

[63] On how nomenclature rules shifted the power between museum and field collectors, and between European and New World collectors, see Christophe Bonneuil, "The Manufacture of Species: Kew Gardens, the Empire and the Standardisation of Taxonomic Practices in Late 19th Century Botany," in *Instruments, Travel and Science: Itineraries of Precision from the Seventeenth to the Twentieth Century*, ed. Marie-Noëlle Bourguet, Christian Licoppe, and Heinz Otto Sibum (New York, 2002), 189–215; Sharon E. Kingsland, *The Evolution of American Ecology, 1890–2000* (Baltimore, 2005), chap. 2; and Endersby, *Imperial Nature* (cit. n. 58), chap. 8.

[64] On the former, see Allen, *Naturalist in Britain*; on the latter, Kohler, *All Creatures*, chap. 2 (Both cit. n. 37), and Mark V. Barrow, *A Passion for Birds: American Ornithology after Audubon* (Princeton, N.J., 1998).

[65] Kohler, *All Creatures* (cit. n. 37), 93.

[66] Cited in Endersby, *Imperial Nature* (cit. n. 58), 39.

[67] J. F. A. Adams, "Is Botany a Suitable Study for Young Men," *Science* 9 (1887): 116–7, on 117. This source is cited in Philip J. Pauly, "Summer Resort and Scientific Discipline: Woods Hole and the Structure of American Biology," in *The American Development of Biology*, ed. Ronald Rainger, Keith R. Benson, and Jane Maienschein (Philadelphia, 1988), 121–50.

[68] On women as a workforce, see Margaret W. Rossiter, *Women Scientists in America: Before Affirmative Action, 1940–1972* (Baltimore, 1995), chap. 3; on women in surveys, Kohler, *All Creatures* (cit. n. 37), 215–20; and on the gendering of biology at the turn of the century, Pauly, "Summer Resort," (cit. n. 67), 129.

[69] Cited in de Chadarevian, *Designs for Life* (cit. n. 15), 89.

positions in natural history museums as directors, like Louis Agassiz and later Mayr at Harvard's Museum of Comparative Zoology, or as curators of specific collections, like Simpson, who was in charge of the American Museum of Natural History's Department of Geology and Paleontology.[70] These positions might have represented professional stability for those who held them but, due to the cultural significance of museums, they also reinforced the association between collectors and leisure activities, especially since the late nineteenth century. The place of collecting has, in large part, defined the identity of the collectors.

As Paula Findlen has so eloquently shown, early modern wonder cabinets, such as that of Ulisse Aldrovandi in Bologna, which contained several thousands of specimens, were hybrid places.[71] Aldrovandi's cabinet, as all others, served many functions. It was a display of his patron's power, a place of civil conversation, and a repository that he used as a basis for the descriptions of animals he published in his numerous natural history books. The leisurely and the scholarly lived side by side. Only in the second half of the nineteenth century did museums begin to make a clearer division between the two. This new "dual arrangement" physically separated the spaces devoted to research and those intended for public displays.[72] Yet, because natural history museums were mainly funded by public monies, philanthropies, and ticket sales, they often emphasized their role as places of education and amusement, rather than research. These museums, with their emblematic dinosaur skeletons, became increasingly associated with other institutions of bourgeois entertainment, such as movie theaters, restaurants, and zoos. This did not raise the scientific stature of the research carried out in rooms behind the museums' lavish dioramas.[73] Even the "research" expeditions of natural history museums were framed as enterprises of exploration and adventure (rather than as purely scientific pursuits) to attract the public attention required to fund these costly endeavors.[74] As a result, collectors were once again trapped by their association with amateurish pursuits. Museums solely devoted to research, such as Berkeley's Museum of Vertebrate Zoology, remained exceptional, far too rare to affect the cultural meaning of sciences dependent on collecting.

Regardless of its association with amateurish activities, within the naturalist community, personal field-collecting experience seems to have been indispensable to making a career.[75] Even naturalists who directed collecting enterprises from their

[70] On the former, see Mary P. Winsor, *Reading the Shape of Nature: Comparative Zoology at the Agassiz Museum*, Science and Its Conceptual Foundations (Chicago, 1991); on the latter, Léo F. Laporte, *George Gaylord Simpson: Paleontologist and Evolutionist* (New York, 2000).

[71] Findlen, *Possessing Nature* (cit. n. 41).

[72] On the dual arrangement, see Lynn K. Nyhart, *Modern Nature: The Rise of the Biological Perspective in Germany* (Chicago, 2009), chap. 6, and Mary P. Winsor, "Museums," in Bowler and Pickstone, *Modern Biological and Earth Sciences* (cit. n. 62), 60–75.

[73] On dioramas at the American Museum of Natural History, see Donna Haraway, "Teddy Bear Patriarchy: Taxidermy in the Garden of Eden, New York City, 1908–1936," *Social Text* 11 (1984–5): 20–64; Stephen C. Quinn, *Windows on Nature: The Great Habitat Dioramas of the American Museum of Natural History* (New York, 2006); Karen Wonders, *Habitat Dioramas: Illusions of Wilderness in Museums of Natural History* (Uppsala, 1993).

[74] On the funding of these expeditions, see Michael Kennedy, "Philanthropy and Science in New York City: The American Museum of Natural History, 1868–1968" (PhD diss., Yale Univ., 1968); on expeditions, Lyle Rexer and Rachel Klein, *American Museum of Natural History: 125 Years of Expedition and Discovery* (New York, 1995).

[75] Ernst Mayr, E. Gorton Linsley, and Robert Leslie Usinger, *Methods and Principles of Systematic Zoology*, McGraw-Hill Publications in the Zoological Sciences (New York, 1953), chap. 4.

desks had been active in collecting specimens in the field at the beginning of their careers. In the nineteenth century, Hooker accompanied a polar expedition to gather the material for his *Botany of the Antarctic Voyage* and traveled to India and the Himalayas before he got a position at Kew Gardens.[76] Similarly, in the twentieth century, Mayr collected thousands of bird skins in Papua New Guinea before he joined the American Museum of Natural History.[77] When these professionals became directors of garden or museum collections, where they relied almost exclusively on existing collections or on other naturalists to gather specimens in the field, they nonetheless claimed an identity of (former) field collectors,[78] unlike theoretical physicists, for example, who would pride themselves on never performing experiments.

THE USES OF COLLECTIONS

The perception that collectors were solely interested in the accumulation of specimens, rather than in the production of knowledge, left most of them without a job, at least in academia. The caricature of the naturalist as a "stamp collector," an expression used at least since the mid-nineteenth century, or the physicist Ernest Rutherford's comment in the first decades of the twentieth century that "all science is either physics or stamp collecting," illustrates the low scientific standing attributed to collecting, especially in the twentieth century.[79] Yet, for most naturalists, collecting was a means to an end with recognizable scientific value: the constitution of a collection that would serve as the basis for their production of biological knowledge.[80]

Even though the institutions hosting collections—museums, gardens, and zoos—have been, since the nineteenth century, places of public enlightenment, moral education, and entertainment, they were (and are) also key places for the production of scientific knowledge. Historians who have worked on natural history museums have emphasized these institutions' role in the collection and display of specimens, but paid less attention to how they were used for the production of knowledge. This circumstance necessitates exploring the role of collections on a more general than specific level, with reference especially to how collection both mirrors and differs from experimentation in practice.

At least since the beginning of the early modern period collections were used to gain insight into the natural world. In wonder cabinets, such as Ferrante Imperator's seventeenth-century cabinet in Naples, the juxtaposition of widely different specimens served to highlight their uniqueness, rarity, or wondrous character.[81] After the

[76] Endersby, *Imperial Nature* (cit. n. 58), chap. 1.

[77] On Mayr's travel, see Jürgen Haffer, *Ornithology, Evolution, and Philosophy: The Life and Science of Ernst Mayr, 1904–2005* (New York, 2007), chap. 2; on his use of collections, Kristin Johnson, "Ernst Mayr, Karl Jordan, and the History of Systematics," *Hist. Sci.* 43 (2005): 1–35.

[78] See, e.g., Mayr's self-characterization in Mayr, Linsley, and Usinger, *Methods and Principles* (cit. n. 75), chap. 4.

[79] On the history of "stamp collecting," see Kristin Johnson, "Natural History as Stamp Collecting: A Brief History," *Arch. Natur. Hist.* 34 (2007): 244–58.

[80] For a good example of the role of collections for systematic work in the twentieth century, see Johnson, "Ernst Mayr" (cit. n. 77).

[81] Findlen, *Possessing Nature* (cit. n. 41); Lorraine Daston and Katharine Park, eds., *Wonders and the Order of Nature, 1150–1750* (Cambridge, Mass., 1998); Robert John Weston Evans and Alexander Marr, *Curiosity and Wonder from the Renaissance to the Enlightenment* (Aldershot, 2006).

collapse of "emblematic natural history,"[82] collections continued to be essential tools for the production of natural knowledge, but in a different epistemological setting. Louis XV's natural history collection, for example, was cataloged by Georges Louis Leclerc Buffon in what eventually became his thirty-six-volume description of the natural objects known to the eighteenth century, the *Histoire naturelle, générale et particulière*.[83] After the Revolution, when the collection was incorporated into the Muséum d'Histoire Naturelle in Paris, it served as a basis for Cuvier's masterful natural histories of quadrupeds and fishes and for his theories of animal anatomy and extinction.[84] In the twentieth century, the elaboration of the evolutionary synthesis by Simpson and Mayr resulted from their extensive examination of the collections of fossils and bird skins of the American Museum of Natural History.

Although these various naturalists' collections were composed according to widely different rules, they all served the same purpose: making systematic comparisons possible by physically juxtaposing different objects. Since the Renaissance at least, collections seem to have worked as material representations of nature, as a "second nature" that could be described, measured, analyzed, and compared in order to generate natural knowledge. These collections can be considered to have been representations of nature, like paintings, because they reflected an intentional perspective, embodied in a narrow selection of natural objects. Furthermore, they were groupings of things as made by collectors, not as found in nature. A collector isolated a thing in nature (say, a bird), stripped it of its relations to its surroundings (the forest), left behind most of its properties (such as being alive), and turned it into a specimen embedded in a new system of relations with other specimens in a collection. Birds could be found in trees, but in collections, there were only specimens. One only needs to think of the indispensable role of taxidermists in preparing specimens for museum conservation to realize how much these are also human artifacts. In this sense, the production of knowledge from collections was no different from the production of knowledge from experiments. The objects of knowledge in the experimental sciences, the "epistemic things" that Hans-Jörg Rheinberger has so productively explored, were not found in nature either; they were made through the human creation of "experimental systems" and the production of controlled phenomena.[85]

Importantly, collections differed from catalogs or repositories of identical things, in that they embodied the idea that the objects they contained were related in some natural (or supranatural) way that the comparative perspective would reveal. After the seventeenth century, natural objects were collected separately from human artifacts, such as scientific instruments, because they were believed to be related in a unique way.[86] Though based on widely different assumptions, Richard Owen's search

[82] William B. Ashworth, "Emblematic Natural History of the Renaissance," in Jardine, Secord, and Spary, *Cultures of Natural History* (cit. n. 7), 17–37.

[83] Buffon, *Histoire naturelle, générale et particulière, avec la description du cabinet du roy* (Paris, 1749).

[84] Dorinda Outram, *Georges Cuvier: Vocation, Science, and Authority in Post-revolutionary France* (Manchester, 1984); Toby A. Appel, *The Cuvier-Geoffroy Debate: French Biology in the Decades before Darwin* (New York, 1987); M. J. S. Rudwick, *Georges Cuvier, Fossil Bones, and Geological Catastrophes: New Translations and Interpretations of the Primary Texts* (Chicago, 1997).

[85] Rheinberger, *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube* (Stanford, Calif., 1997); Ian Hacking, "The Self-Vindication of the Laboratory Sciences," in *Science as Practice and Culture*, ed. Andrew Pickering (Chicago, 1992), 29–64.

[86] Pomian, *Collectors and Curiosities* (cit. n. 60).

for an archetype and the post-Darwinian search for common descent both proceeded through the comparison of specimens in collections that were believed to share common properties.[87] The notion of homology served as a guiding principle to organize collections and make comparisons. This helps clarify the conceptual limits of the category of *collection*. Collections assembled things that were believed to be related in nature, not just in a researcher's mind. This also justifies why they can be considered representations, because they bore this kind of epistemological relationship to the natural world.

Collections can also be conceptualized as "relational systems." Unlike experimental systems, which offer the possibility to manipulate and create differences in a single object, relational systems make the systematic comparison of many objects possible. Natural history collections were composed not only of individual things, but also of all the many relations among the things they contained. As a result, their epistemic potential was understood to grow exponentially with their size and was driven by the ideal of "completedness." As Buffon put it, "At each sight, not only does one gain a real knowledge of the object considered, but furthermore one discovers the relationships it can have with those around it."[88]

Comparisons such as those made by Buffon and other naturalists required more than the spatial concentration of things. Scientific collections, unlike many others, made the comparison of apples and oranges (or of bacteria and elephants) possible by performing two operations: an ontological reduction and a formal standardization. Elements in a collection were each reduced to a common set of properties (bones or skins) and then were formatted identically (as mounted specimens). Taken alone, these operations might seem trivial, yet together they potentiated the epistemic function of collections. They made the production of general knowledge possible through the comparison of numerous items (a logically invalid, but practically valuable, form of induction). For example, from the comparison of numerous bird specimens with one another, researchers drew general conclusions about classes of things—types of birds—and about their structure, function, and history.

Experimentalists achieved the aim of producing general knowledge in a different way: they paired the use of a few carefully selected (mainly for practical reasons) "exemplary" model systems with the assumption that these systems were representative across broader classes of things. The exemplary perspective has been as important for experimentalists as the comparative perspective has been for naturalists.[89] From Claude Bernard's use of dogs as models of human physiology to geneticists' use of fruit flies as models of genetic transmission, the growth of the experimental life sciences depended on the development of model organisms.[90] These provided the laboratory researchers' "second nature." Experimentalists firmly believed that the knowledge produced with these few model species was of universal validity. The French molecular biologists Jacques Monod and François Jacob (paraphrasing Albert Kluyver) put it best in 1961: "[What is] true of *E. coli* must also be true of

[87] Nicolaas A. Rupke, *Richard Owen: Biology without Darwin* (Chicago, 2009).

[88] Buffon, *Histoire naturelle* (cit. n. 83), 4; translation mine.

[89] For a broader discussion of this distinction, see Strasser and de Chadarevian, "Comparative and the Exemplary" (cit. n. 21).

[90] Frederic Lawrence Holmes, *Claude Bernard and Animal Chemistry: The Emergence of a Scientist* (Cambridge, 1974); Robert E. Kohler, *Lords of the Fly:* Drosophila *Genetics and the Experimental Life* (Chicago, 1994).

Elephants."[91] Needless to say, Monod and Jacob never brought an elephant into the laboratory to check this assumption.

Experimental systems and relational systems both produced universal knowledge, via not only an abstract intellectual operation, but also a material transformation. As Bruno Latour has suggested, knowledge produced in a laboratory is made universal by extending the conditions of the laboratory to the outside world, including to other laboratories.[92] Facts produced experimentally in a laboratory somewhere can be replicated in laboratories anywhere, not only because of underlying regularities in nature, but also because laboratories have been made almost identical to one another through the standardization of instruments, protocols, and skills. Laboratories can thus be pictured as "centrifugal places": facts travel outward from their initial site of production. Collections, by contrast, can be pictured as "centripetal places": they concentrate objects often otherwise dispersed around the world (such as plants and animals) and partially standardize them in order to make them more easily comparable. When objects become accessible in a single place, in a single format, they can be arranged to make similarities, differences, and patterns apparent to the eye of a single human investigator; collections concentrate the world, making it accessible to the limited human field of view. As Buffon put it in 1749, "The more you see, the more you know."[93]

This brief overview of collecting practices, the identity of collectors, and the epistemic uses of collections serves as a backdrop to my historical reconstruction of the surprising development of collections of biological things and data in the twentieth century. Later collectors faced some of the same challenges as their predecessors, but in a very different context. As briefly outlined in the introduction of this article, historians of the life sciences have explored in great detail the rhetorical and institutional battles between experimentalists ("laboratory men") and naturalists ("museum men") at the turn of the twentieth century and between molecular biologists and evolutionary biologists at midcentury.[94] Historians of the life sciences have also investigated how naturalists responded to the dominance of experimentalism and the transformations that took place within natural history.[95] But what historians have not yet explored is the fact that the same debates took place not just between experimentalists and naturalists or within natural history, but also between experimentalists and within the experimental life sciences themselves. The stellar rise of the experimental life sciences in the twentieth century obscured the fact that their success was not

[91] Monod and Jacob, "General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation," *Cold Spring Harbor Symposia on Quantitative Biology* 21 (1961): 389–401.

[92] Latour, "Give Me a Laboratory and I Will Raise the World," in *The Science Studies Reader*, ed. Mario Biagioli (New York, 1999), 258–75.

[93] Buffon, *Histoire naturelle* (cit. n. 83).

[94] Allen, "Morphology and Twentieth-Century Biology" (cit. n. 13); Dietrich, "Paradox and Persuasion"; Hagen, "Naturalists" (Both cit. n. 16).

[95] Keith Vernon, "Desperately Seeking Status: Evolutionary Systematics and the Taxonomists' Search for Respectability, 1940–60," *Brit. J. Hist. Sci.* 26 (1993): 207–27; Joel B. Hagen, "Experimental Taxonomy, 1920–1950: The Impact of Cytology, Ecology, and Genetics on the Ideas of Biological Classification" (PhD diss., Oregon State Univ., 1984); Hagen, "Experimentalists and Naturalists in 20th-Century Botany—Experimental Taxonomy, 1920–1950," *J. Hist. Biol.* 17 (1984): 249–70; Bruno J. Strasser, "Laboratories, Museums, and the Comparative Perspective: Alan A. Boyden's Serological Taxonomy, 1925–1962," *Hist. Stud. Nat. Sci.* 40 (2010): 149–82.

necessarily the result of experimental practices, but emerged also, as I argue, from practices centered on collections.[96]

COLLECTING EXPERIMENTS IN THE TWENTIETH CENTURY

As an initial step in considering the place of collecting and collections in the twentieth-century experimental life sciences, let us reexamine the paradigmatic example of the experimentalists' triumph: molecular biology. The stories of the greatest successes of molecular biology—determining the three-dimensional structure of proteins, understanding the structural basis of their function, and deciphering the genetic code—have all been told as having resulted from experimental virtuosity (generally leading to Nobel Prizes). In the last case, the narrative is made all the more poignant by the success of two relatively unknown researchers in 1962 in cracking the code experimentally after a number of great minds, including Francis Crick, had tried unsuccessfully for years to find a solution theoretically. But as Soraya de Chadarevian and I have shown elsewhere, these achievements were due to a combination of experimental and comparative approaches, not to experimental breakthroughs alone.[97] Frederick Sanger, who determined for the first time the sequence of a protein (insulin isolated from an ox), was at a loss to identify which part of the molecule played a significant role for its biochemical function—that is, at a loss until he sequenced insulin molecules from several other species, compared the sequences, and identified specific regions that had remained constant throughout evolution. Similarly, to understand the structural basis of the hemoglobin molecule's function, Max Perutz relied on an extensive collection of hemoglobin molecule variants, which he compared systematically.[98] Finally, the first codon of the genetic code was determined thanks to an ingenious experiment, but in the determination of the remaining sixty-three codons, collections of protein sequences from various organisms proved to be a tremendous asset.

These examples were not isolated cases in the history of molecular biology. Scientific recognition and public visibility have usually gone to the authors of audacious experiments performed on "exemplary" cases in model organisms and model systems. But these achievements were often made possible by the comparison of experimental data from a much wider range of organisms and systems. The results obtained on these other systems were not mere repetitions of the initial finding; instead, the accumulation of results opened up new epistemic possibilities—namely, systematic comparisons. In the second half of the twentieth century, public collections of experimental data became increasingly common, providing researchers with the material they needed for making such comparative studies.

In the scholarly discussion that followed the publication of Allen's history of the twentieth-century life sciences, one point became clear. Even though experimentalism grew tremendously in this period and captured most of the attention, natural

[96] For an examination of this claim for the history of molecular biology, see Strasser and de Chadarevian, "Comparative and the Exemplary" (cit. n. 21).
[97] Ibid.
[98] Soraya de Chadarevian, "Following Molecules: Haemoglobin between the Clinic and the Laboratory," in *Molecularizing Biology and Medicine: New Practices and Alliances, 1910s–1970s*, ed. de Chadarevian and Harmke Kamminga (Amsterdam, 1998), 171–201.

history did not disappear. Lynn K. Nyhart argued that natural history might have become secondary to the experimental life sciences, but it kept growing with the general expansion of biology, while Keith R. Benson claimed that natural history remained "alive and well," although "primarily within museums."[99] In the twentieth century, collecting in the field (e.g., to provide specimens for natural history museums or data for ecological studies) remained an important activity, and naturalist collections, of specimens and data alike, continued to grow—and historians of science have hardly begun to explore this area.

I go one step beyond this argument that natural history (with natural history collecting) was "alive and well" in the twentieth century, to contend that collecting was also an essential practice for the experimental sciences in the twentieth century and that they, too, relied on collections for the production of knowledge. As I have shown in several prior papers, numerous collections of data about the structure and function of molecules began to be assembled in the 1960s.[100] Almost half a century later, these collections not only still exist, but they have become indispensable tools for most laboratory researchers. Before discussing the historiographic benefits of considering these collections in the same light as the naturalists' collections, a brief overview of their development is in order.

The development of data collections in the twentieth century resulted from an increasing rate in the production of data, the perception of an "information overload," the intellectual opportunities offered by systematic comparisons of data, and the power offered by computers and eventually computer networks to conduct these on a large scale. The accumulation of data not only made the creation of collections possible, it made them increasingly useful. At the same time, these collections often facilitated experiments that produced even more data.[101]

The *Atlas of Protein Sequences and Structure*, for example, a collection of data about protein sequences, was first published in 1965 by the physical chemist Margaret O. Dayhoff.[102] She justified the creation of her collection with this statement: "There is a tremendous amount of information regarding evolutionary history and biochemical function implicit in each sequence and the number of known sequences is growing explosively." She felt that it was "important to collect this significant information, correlate it into a unified whole and interpret it."[103] Indeed, starting in the 1960s, the pace of protein sequencing was becoming "fast and furious."[104] In 1968, an editorial in *Science* made the point that the determination of protein sequences

[99] Nyhart, "Natural History and the 'New' Biology," in Jardine, Secord, and Spary, *Cultures of Natural History* (cit. n. 7), 426–43, on 422; Benson, "From Museum Research to Laboratory Research: The Transformation of Natural History into Academic Biology," in Rainger, Benson, and Maienschein, *American Development of Biology* (cit. n. 67), 49–83, on 77.

[100] See esp. Bruno J. Strasser, "Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequences and Structure, 1954–1965," *J. Hist. Biol.* 43 (2010): 623–60, and Strasser, "The Experimenter's Museum: GenBank, Natural History, and the Moral Economies of Biomedicine," *Isis* 102 (2011): 60–96.

[101] For an example of the same dynamic in early modern collections, see Isabelle Charmantier and Staffan Müller-Wille, "Natural History and Information Overload: The Case of Linnaeus," *Stud. Hist. Phil. Biol. Biomed. Sci.* 43 (2012): 4–15.

[102] Strasser, "Collecting, Comparing, and Computing" (cit. n. 100).

[103] Dayhoff to Carl Berkley, 27 February 1967, National Biomedical Research Foundation Archives, Georgetown University, Washington, D.C. (hereafter, NBRF Archives).

[104] "Proteins: Yet More Sequences," *Nature* 224 (1969): 313.

was "one of the most important activities today."[105] The "explosion" in sequence data that Dayhoff and others observed resulted from several factors, including the development in 1967 of Pehr Edman's Sequenator, a rapid and efficient automatic protein sequencer. The availability of this machine emboldened researchers to take on the challenge of larger and more complex proteins. The rising interest in molecular evolution also led a number of researchers to sequence proteins from ever more diverse species. The *Atlas* itself further facilitated these sequencing efforts by offering researchers a number of homologous sequences with which they could compare their partial experimental results, thus contributing to the growth of sequence data that it was supposed to tame.

Similarly, the creation in 1973 of the Protein Data Bank (PDB), a collection of data about the three-dimensional structure of proteins, followed the announcement at the Cold Spring Harbor Symposia on Quantitative Biology two years earlier that several new protein structures had been solved thanks to improved crystallographic methods.[106] Expecting a rapid growth in the number of protein structures solved experimentally and counting on the possibilities for comparison offered by a collection, the crystallographers Helen M. Berman, Edgar F. Meyer, and Walter C. Hamilton established the PDB at Brookhaven National Laboratory and began to distribute the data describing the structure of proteins. Like the *Atlas*, the PDB greatly facilitated the determination of new protein structures, thus adding to the growth of crystallographic data.

A decade later, the European Molecular Biology Laboratory and the NIH sponsored the creation of DNA sequence collections.[107] Again, the creation of these collections was prompted by breakthroughs in the methods to produce data. In this case, two new methods to sequence DNA were developed in 1977 that led to an exponential increase in the amount of sequence data and plans to organize them in a collection. Yet, as two molecular researchers put it shortly afterward, "the rate limiting step in the process of nucleic acid sequencing is now shifting from data acquisition towards the organization and analysis of that data."[108] When the European Molecular Biology Laboratory's DNA sequence library and the NIH's GenBank became publicly available in 1982, they too fueled the explosion of data.

These collections and the many others that were created in the same period grew rapidly in size and popularity. Only seven years after its first edition, the *Atlas* had grown tenfold, and five years later, it was among the fifty most cited scientific items of all time.[109] Similarly, GenBank, today's largest collection of biological information, has grown to contain over one hundred billion As, Ts, Gs, and Cs, which amounts to as many letters as are found in two thousand copies of Buffon's thirty-six-volume *Histoire naturelle*. In 2011, more than twenty thousand computers connected directly to GenBank every day, indicating an even larger number of actual users.[110]

[105] Philip H. Abelson, "Amino Acid Sequence in Proteins," *Science* 160 (1968): 951.

[106] Helen M. Berman, "The Protein Data Bank: A Historical Perspective," *Acta Crystallographica A* 64 (2008): 88–95.

[107] Strasser, "Experimenter's Museum" (cit. n. 100).

[108] Thomas R. Gingeras and Richard J. Roberts, "Steps toward Computer Analysis of Nucleotide Sequences," *Science* 209 (1980): 1322–8.

[109] Margaret O. Dayhoff to Donald DeVincenzi, 10 June 1980, NBRF Archives.

[110] Dennis Benson (National Center for Biotechnology Information), personal communication to the author, 3 October 2011.

The key point is that these collections are part and parcel of experimental research carried out within laboratories. One researcher reported to Dayhoff that the *Atlas* "is the most heavily used book in our lab," while another confessed, "We use your book like a bible!" (a strange comparison indeed).[111]

If we want to take seriously the resemblances between these more recent databases and earlier collections, we need to ask some of the same questions about the collecting practices that led to their establishment and that support their continued growth as we asked earlier about older practices. How was collecting performed? Where did the items in the collections come from? Who were the collectors? Why did they collect? What were the epistemic and social rewards of their collecting enterprises? And finally, how were the collections used for the production of knowledge?

TAKING SPECIMENS FROM THE FIELD TO THE LABORATORY

The history of the experimental life sciences has been told from the vantage point of the few organisms—*Drosophila*, corn, and mice—that have served as model organisms.[112] Yet experimentalists produced data about a much broader range of species, including wild ones, such as the badger, bison, fox, green monkey, guinea pig, llama, mink, red deer, and reindeer. How did these organisms of the field become laboratory objects? The animal materials used in modern protein research came from sources both close to home and far away; they were obtained from laboratory researchers occasionally venturing into the field and from professional naturalists, animal dealers, and zookeepers. In their variation and extent, these modes of collecting are suggestive in many ways of early natural history networks.

In most cases, researchers investigating proteins obtained their material from local slaughterhouses where they could purchase large amounts of tissues for a low price, most often organs that were not sold for human consumption. Proteins were then carefully extracted and purified in the laboratory. As a result of this particular economy, many studies were conducted on cows, pigs, horses, and chickens. Biochemists purified cytochrome *c* proteins, for example, from beef, using a few "freshly minced" heart muscles.[113] The meatpacking industry also provided material for scientists. For example, the Chemical Research and Development Department of Armour and Company (best known in the postwar United States for its hot dogs with "open fire flavor") purified ribonuclease, lysozyme, and other proteins from bovine pancreases and put them up for sale to researchers.[114] The whaling industry was another source; it provided sperm whale meat for Perutz's studies of hemoglobin in Cambridge and supplied other studies of insulin carried out in Japan, where the whaling industry was well developed.[115]

Human samples came from equally diverse and extended networks. Pathological hemoglobins in humans, for example, were provided by clinics in regions where the

[111] Allen B. Edmundson to Robert S. Ledley, 25 November 1969, and Oliver Smithies to Winona Barker, 5 October 1970, NBRF Archives.

[112] See, e.g., Jim Endersby, *A Guinea Pig's History of Biology* (Cambridge, 2007).

[113] David Keilin, "Preparation of Pure Cytochrome *c* from Heart Muscle and Some of Its Properties," *Proceedings of the Royal Society of London B—Biological Sciences* 122 (1937): 298–308.

[114] "Science Exhibition," *Science* 106 (1947): 567–75.

[115] Georgina Ferry, *Max Perutz and the Secret of Life* (New York, 2007).

prevalence of certain diseases was high. For his investigations of sickle-cell anemia hemoglobin, most prevalent in the United States among African Americans, Linus Pauling secured a blood supply from a clinician in New Orleans to use in his laboratory at Caltech.[116] In Cambridge, England, Vernon Ingram relied on sickle-cell anemia blood brought by Anthony C. Allison from Kenya.[117] Later, Ingram explored the molecular differences in hemoglobin from patients with many different pathological conditions. In this case, the blood samples were taken from the blood collection that the clinician Hermann Lehmann had established in Cambridge from his trips in several African countries.[118]

The supply of biological material from wild animals posed a greater challenge to laboratory workers. As for earlier naturalist collections, local environments played a defining role. Indeed, protein sequences from deer were determined in a laboratory in Stockholm, those from camels in Udaipur, and those from rattlesnakes in Los Angeles.[119] Unlike museum naturalists, most laboratory biochemists had no prior experience of field collecting, and they gathered material from the immediate surroundings of their laboratories. The chemists Margareta and Birger Blombäck at the Karolinska Institute in Stockholm, for example, were leading researchers on the molecular basis of blood coagulation in the 1960s, but their only use for the outdoors had been recreational. For their studies on the mechanisms of coagulation, however, they needed and secured blood samples from a wide range of organisms, beginning with domestic animals, and moving later to wild ones. Their interest in samples from the latter turned them into field-workers. Together with a visitor from the United States, the biochemist Russell F. Doolittle, they flew to northern Sweden for the annual reindeer hunt, where "a Laplander and his lasso" captured a few specimens from which blood was drawn.[120] They also traveled Sweden's northern islands to hunt seals whose blood was then investigated in Stockholm.[121] The problem of storing biological samples at a freezing temperature, which had stymied so many field-collecting expeditions for blood in Africa and Central America, was easily solved in Lapland, with "nature providing excellent refrigeration."[122]

Like early field collectors of natural history, the Blombäcks were interested in expanding their collection whenever the opportunity arose to do so. In 1963, they had moved temporarily to Australia and seized this chance to gather blood from different species of kangaroos and sharks that were readily accessible in this new environment. That same year, they extended their interests in fibrinopeptide variation to human populations, again because a new diversity of types was available in their new surroundings. Margareta Blombäck wrote enthusiastically that they had gathered "blood

[116] Lily E. Kay, *The Molecular Vision of Life: Caltech, the Rockefeller Foundation and the Rise of the New Biology* (New York, 1993).

[117] Ingram, "Sickle-Cell Anemia Hemoglobin: The Molecular Biology of the First 'Molecular Disease'—the Crucial Importance of Serendipity," *Genetics* 167 (2004): 1–7.

[118] De Chadarevian, "Following Molecules" (cit. n. 98).

[119] Margaret O. Dayhoff, *Atlas of Protein Sequence and Structure* (Silver Spring, Md., 1972).

[120] John F. Henahan, "Dr. Doolittle—Making Big Changes in Small Steps," *Chemical and Engineering News*, February 9, 1970, 22–32, on 23.

[121] Margareta Blombäck, personal communication to the author, 19 May 2010; Blombäck, "Thrombosis and Haemostasis Research: Stimulating, Hard Work and Fun," *Thrombosis and Haemostasis* 98 (2007): 8–15.

[122] Henahan, "Dr. Doolittle" (cit. n. 120), 23.

from different [human] races, as pure as they possibly can be, such as Maoris (New Zealand), New Guinea natives, East Africans and Australian Negros" and that they had started "a new field of biochemical anthropology."[123]

In addition to field collecting, the Blombäcks, like many other biochemists and naturalists, relied on gifts from individual colleagues around the world who had access to local species. The method had its limitations, mainly because the regions hosting the most exotic species also had the least number of laboratories. When he was unable to obtain blood from a rare species of monkey living on just a few Southeast Asian islands for his hemoglobin studies of primates, the anthropologist John Buettner-Janusch, at Yale University, complained, "We have not yet been able to beg, borrow, or steal a sample of Tarsius hemoglobin."[124] Most researchers adopted the same strategy as Doolittle, who worked in San Diego and relied extensively on the exceptionally rich animal collection present in its public zoo. Marine stations, such as the Marine Biological Laboratory in Woods Hole, Massachusetts, and the marine station of the Collège de France in Concarneau, Brittany, were used as sources for aquatic animals.[125]

Examined from the perspective of where and how specimens were obtained, laboratory research on the molecular basis of protein function begins to resemble the collecting endeavors usually associated with natural history. The same logic of place prevailed: collectors first assembled local species, and then more distant ones, in a quest to have the broadest number of species represented. The same logic of assembling and using a collection prevailed too: these researchers brought the different protein sequences into a common format to make them comparable, performed systematic comparisons, and drew general conclusions about the structure, function, and history of these proteins. This story of collecting and comparing begins to differ from the story that attributes the successes of the molecular life sciences to experimental virtuosity and single model organisms.

THE COLLECTORS' STANDING IN SCIENCE

The vast majority of collectors of experimental data were not naturalists, although a few naturalists collected experimental data in the twentieth century, such as Alan Boyden, in his Serological Museum at Rutgers University, or Charles Sibley, in his collection of bird DNA at Yale University. Most were trained experimental scientists, and many had doctoral degrees, in fields such as physical chemistry or crystallography. Their experiences, however, in some ways paralleled those of naturalist collectors of an earlier period. Perhaps most obviously, data collectors, such as Dayhoff (of the *Atlas*), Berman (of the PDB), and Olga Kennard (of the Cambridge Crystallographic Data Centre), were often women and they relied on extensive female staffs, not unlike many earlier collecting enterprises. Certainly, several men devoted their (late) career to collecting,[126] such as Walter Goad with GenBank, but women

[123] M. Blombäck to her parents, 19 September 1963, Margareta Blombäck personal archives.

[124] Buettner-Janusch and R. L. Hill, "Molecules and Monkeys," *Science* 147 (1965): 836–42.

[125] Doolittle, "Characterization of Lamprey Fibrinopeptides," *Biochemical Journal* 94 (1965): 742–50; R. Acher, J. Chauvet, and M. T. Chauvet, "Phylogeny of the Neurohypophysial Hormones," *Nature* 216 (1967): 1037–8.

[126] In another field, see Michael D. Gordin, "Beilstein Unbound: The Pedagogical Unraveling of a Man and His Handbuch," in *Pedagogy and the Practice of Science: Historical and Contemporary Perspectives*, ed. David Kaiser (Cambridge, Mass., 2005), 11–39.

made up an unusually high proportion of collectors, especially in comparison to their marginality in the fields from which the data were being collected. In addition, their professional identity within the experimentalist community was very unstable. Although they were trained in the experimental sciences and worked with experimental data, their collecting work was generally considered not to be of a scientific nature. Experimentalists writing to Dayhoff, for example, addressed her as a "compiler," an "editor," or a "librarian," none of which was a very enviable status to aspire to for a scientist.[127]

As I have discussed elsewhere, viewing Dayhoff's work as part of the collecting tradition, resting on different epistemic, social, and cultural norms than the experimental sciences, helps us understand the difficulties in the development of her professional career.[128] She was denied membership in the American Society of Experimental Biologists because, according to one of its members, the "compilation of the Atlas of Protein Sequence and Structure" could not be considered her "own research."[129] In other words, the problem with Dayhoff's collection-based work was not only that it was not experimental, but also that it was collective, and thus did not fit into the highly individual reward ethos of the experimental sciences. Similarly, the NIH remained reluctant until the 1980s to fund data collections, because they did not fit a grant system geared toward individual experimental research. In 1981, after the NIH had turned down one of her grant requests for the *Atlas*, Dayhoff lamented, as she had at other times, "Databases do not inspire excitement."[130]

Most experimentalists considered the work of data collectors mundane, clerical, or even trivial. They overlooked the data collectors' wide range of expertise. First, the experimental data gathered by the collectors were often plagued with errors, many due to simple transcription mistakes by the authors and publishers of the data. It took a precise understanding of the nature of proteins and of the methods (biochemical or crystallographic) that had been used to produce the data to spot the possible errors and resolve them with the authors. Since the data collections were intended to be not mere repositories, but tools for the production of knowledge, the collectors organized the data in ways that would be most productive epistemically. This task required collectors to understand how the data could be used in research. For example, the protein sequences contained in the *Atlas* were aligned in order to highlight their similarities and differences. This was no trivial task. Since sequences were not identical, "gaps" were inserted in them to optimize the extent of the alignment between a given two, making an implicit assumption about their evolution. Collectors also used the taxonomies they created to organize their data collections. After 1974, for example, the *Atlas* was structured around "superfamilies," a concept introduced by Dayhoff, after an extensive study of all sequences present in the *Atlas*.[131]

Just as the long history of collecting pointed to the importance of mobilizing large networks of collectors in the making of collections, so too does the history of recent collecting call attention to the importance of numerous "field" collectors and to the

[127] E.g., Richard Synge to "Compilers," 7 April 1966, NBRF Archives.
[128] Bruno J. Strasser, "Collecting and Experimenting: The Moral Economies of Biological Research, 1960s–1980s," *Preprints of the Max Planck Institute for the History of Science* 310 (2006): 105–23; Strasser, "Collecting, Comparing, and Computing" and "Experimenter's Museum " (Both cit. n. 100).
[129] John T. Edsall to Dayhoff, 4 November 1969, NBRF Archives.
[130] Dayhoff to D. M. Moore, 24 September 1981, NBRF Archives.
[131] Dayhoff, "Computer Analysis of Protein Sequences," *Federation Proceedings* 33 (1974): 2314–6.

moral economies on which these coordinated collecting enterprises were based. In 1965, the first edition of the *Atlas* included contributions from just over 150 researchers; in 2011 more than 20,000 scientists submitted DNA sequences to GenBank. How were these individuals brought to participate in the collection of data?

Dayhoff encountered great difficulties in obtaining sequence data from researchers. In an earlier paper, I have shown how the failure of Dayhoff's efforts at securing the participation of individual experimentalists was a result of divergent moral economies.[132] Experimentalists had a strong sense of ownership in the data they produced and were unwilling to share them openly for others to use. The fact that Dayhoff copyrighted the data she received and used them for her own research was deemed unacceptable to many experimentalists who had spent months or even years producing these data. Many wanted to retain a symbolic form of ownership over them, in addition to being able to exploit them further.

But Dayhoff's difficulties in obtaining sequence data from researchers paled in comparison to the resistance encountered by collectors of crystallographic data.[133] Those who set up the PDB to collect all known protein structures were often unable, despite repeated calls and pleas, to secure crystallographic data from individual researchers. There was a common agreement that data supporting published interpretations should be publicly available. But the very nature of what constituted "data" was hotly debated. Researchers were most reluctant to share "raw" data as opposed to "processed" data, or "results," arguing that raw data belonged to the inner workings of a laboratory. Others argued that the failure of many crystallographers to make their data public, either in print or electronically through the PDB, undermined the very notion of a "publication." As one crystallographer put it, "Results without data are unproven, and interpretations without results are hearsay." After noting that in three-quarters of the cases raw data were unavailable in publications of certain molecular structures, he concluded that they were "not really published at all, in the literal sense of making the information public." In macromolecular crystallography, he noted, "a custom of non-publication" had been "allowed to grow from an idiosyncrasy, to an inconvenience, to an outright scandal."[134]

By the end of the 1980s, after years of intense negotiations, several crystallographers succeeded in convincing scientific journals to enforce a mandatory submission policy.[135] Only those papers for which the data had been deposited in the PDB would be cleared for publication. At the same time, the managers of GenBank, the collection of nucleic acid sequences, arrived at similar arrangements with journal editors, effectively solving the problem of data collection. What appeared from the 1990s to be a spontaneous communal effort to produce and share data was, in fact, the result of a hard-fought battle that succeeded in balancing the risks and benefits of sharing scientific data.

The struggles encountered by collection managers in securing data from individual researchers invite comparisons with naturalists' collections, such as those in museums of natural history. Naturalists relied and continue to rely on large numbers of

[132] Strasser, "Collecting and Experimenting" (cit. n. 128); see also Strasser, "Experimenter's Museum" (cit. n. 100).
[133] Marcia Barinaga, "The Missing Crystallography Data," *Science* 245 (1989): 1179–81.
[134] Richard E. Dickerson to president of the American Crystallographic Association, July 1987, Protein Data Bank Archives, Rutgers University, New Brunswick, N.J.
[135] J. L. Sussman, "Protein Data Bank Deposits," *Science* 282 (1998): 1993.

amateurs (including many women in botany) to gather data and specimens. These field collectors were often content to give their findings to a local natural history museum, and they felt honored to be mentioned in a scientific publication. Their exclusion from authorship, either in the naming of species or in the publication of taxonomic descriptions, for example, was made easier by the difference in status, and often the gender divide, between professionals and amateurs. These convenient arrangements, however, were not available to modern data collectors: the data they gathered had been produced by scores of professional experimenters who intended to be fully credited for any interpretive work that was based on their data. Furthermore, experimenters believed that the first interpretation of the data belonged to them. For example, in 2002, a genomic researcher from the Marine Biological Laboratory in Woods Hole complained that he had been "scooped with his own data," in an episode that a *Nature* writer called the "latest in a string of clashes between those who collect and those who interpret data."[136]

There was increased acknowledgment that data producers and data analyzers filled different professional niches (as did field collectors and museum taxonomists), but what was the proper place of the data curators and those who assembled data collections? Curators were legitimate figures in natural history (Mayr and Simpson were curators), but had no comparable position in the experimental sciences. If the data in their collections were to be made public and they were denied privileged access to these data (through which they could make scientific contributions), their role would be reduced to that of infrastructure managers, not scientists. In the nineteenth century, Augustin Pyramus de Candolle considered the ownership of an herbarium to be a prerequisite to being a botanist;[137] in twentieth-century molecular sciences, managing a data collection almost prevented one from being a scientist. Unsurprisingly, data collectors (the majority, PhD-carrying scientists) have been dissatisfied by this lack of professional recognition. Some have been able to derive their professional legitimacy from publishing original methods of data analysis; for example, David Lipman, head of GenBank since 1989, has contributed to the development of BLAST (the Basic Local Alignment Search Tool), the most widely used algorithm to compare sequences.[138] However, the professionalization of the role of database curator and manager, aligned with the development of similar professional roles in the digital information and library sciences,[139] has produced an ambiguous legitimacy for researchers in the natural sciences who work with databases of experimental knowledge. As a recent paper put it, database curators "dread the immortal cocktail party question 'So, what do you do?'"[140]

[136] The Woods Hole group had determined the sequence of a bacterium and made the data available online, only to see another group publish an evolutionary interpretation of these data before they were able to propose one themselves. E. Marshall, "Data Sharing—DNA Sequencer Protests Being Scooped with His Own Data," *Science* 295 (2002): 1206–7, on 1206.

[137] Peter F. Stevens, *The Development of Biological Systematics: Antoine-Laurent de Jussieu, Nature, and the Natural System* (New York, 1994).

[138] S. F. Altschul et al., "Basic Local Alignment Search Tool," *Journal of Molecular Biology* 215 (1990): 403–10.

[139] E.g., *Database: The Journal of Biological Databases and Curation* was launched in 2009, two years after the *International Journal of Digital Curation*.

[140] Kyle Burkhardt, Bohdan Schneider, and Jeramia Ory, "A Biocurator Perspective: Annotation at the Research Collaboratory for Structural Bioinformatics Protein Data Bank," *PLoS Computational Biology* 2 (2006): 1186–9, on 1186.

HOW ARE DATABASES USED?

Since their inception in the 1960s, electronic databases, like earlier collections, have been used for generating knowledge on a variety of topics, but always through comparison. Comparison has been the key epistemic practice in producing knowledge about the relationship between form and function, the history of organisms and their parts, and the systematic relationships between organisms. Like Vicq d'Azyr, Cuvier, and other comparative anatomists a century earlier, biochemists have assembled collections of structures that have then served as the basis for their comparative studies.[141] The American biochemist Christian B. Anfinsen, in his 1959 book, *The Molecular Basis of Evolution*, did much to popularize the comparative approach among protein researchers, as well as the idea that similarities in sequence would indicate "the minimum structure which is essential for biological function."[142] These and other biochemists have relied on the diversity of nature—as accessed through their collections—to gain insights into the relationship between the structure and the function of proteins.

The reconstruction of the history of life has long relied on the collection of existing and extinct specimens. Unsurprisingly, phylogenetic research became one of the key uses of molecular databases, such as the *Atlas* and GenBank, following the development of methods in molecular evolution. Dayhoff, for example, pioneered methods to infer phylogenetic distances from numbers of differences between protein sequences.[143]

Databases have also been widely used to elaborate taxonomies of their elements, whether protein structures or DNA sequences. As mentioned previously, Dayhoff proposed the concept of "protein superfamilies," a clear analogy to the taxonomy of species, to designate groups of proteins that shared a similar structure and that had evolved from a unique protein. She derived this concept from the comparison of the data present in her collection and then used it to reorganize the collection according to these categories, much in the same way that natural history collections were (and are) structured by their taxonomic rank. Similarly, the PDB has been used to classify proteins according to their three-dimensional shape. The comparison of shapes, unlike that of sequences, does not lend itself so easily to numerical approaches. Thus, those who have attempted to develop taxonomies of protein structures have resorted to strategies very similar to those used by naturalists in comparing specimens.

Among the many researchers who have adopted the comparative approach in classifying protein structures, the case of Jane S. Richardson is particularly illuminating, as an example not only of this approach but also of the recognition among some scientists of the alignment of their practices with those of natural history collecting and comparing. Without a graduate degree in science (she had a master's in philosophy and had taken some courses in plant taxonomy and evolution at Harvard), she joined a chemistry laboratory at MIT as a technician.[144] She became interested in pro-

[141] On d'Azyr, see Stéphane Schmitt, "From Physiology to Classification: Comparative Anatomy and Vicq D'Azyr's Plan of Reform for Life Sciences and Medicine (1774–1794)," *Sci. Context* 22 (2009): 145–93.

[142] Anfinsen, *The Molecular Basis of Evolution* (New York, 1959), 143.

[143] Joseph Felsenstein, *Inferring Phylogenies* (Sunderland, Mass., 2004), chap. 10.

[144] S. Bahar, "Ribbon Diagrams and Protein Taxonomy: A Profile of Jane S. Richardson," *Biological Physicist* 4, no. 3 (2004): 5–8.

tein structures and elucidated several of them, before focusing on their classification. In the mid-1970s, she started to systematically survey all known protein structures, visually identifying different patterns. She used these patterns, which she compared to geometric motifs common on Greek and American Indian weaving and pottery, as a basis for her classification, which made the cover of *Nature* in 1977.[145] Her work culminated a few years later in an almost two-hundred-page review of the "anatomy and taxonomy of protein structure," which made extensive use of the data contained in the PDB.[146] She grouped all known proteins into classes according to their structures and provided simplified representations of each that would make their common features more apparent. For the same reason, she conceived a new representation of a structural pattern (the beta-sheet) that soon became a standard in protein science.

Richardson explicitly acknowledged how much her comparative approach derived from natural history:

> The vast accumulation of information about protein structures provides a fresh opportunity to do descriptive natural history, as though we had been presented with the tropical jungles of a totally new planet. It is in the spirit of this new natural history that we will attempt to investigate the anatomy and taxonomy of protein structures.[147]

Richardson confessed her "love of complex primary data and what is essentially a new kind of natural history."[148] The objects that Richardson classified might have been the product of experimental virtuosity, but the ways in which she approached them were clearly in line with the natural history tradition. Furthermore, her approach to taxonomy, like that of traditional naturalists, relied not only on the visual inspection of structure, but also on an intimate, personal, and even intuitive grasp of similarities. She later explained that she believed in the importance of

> exhaustively *looking*, in detail, at each beautifully quirky and illuminating piece of data with a receptive mind and eye, as opposed to the more masculine strategy of framing an initial hypothesis, writing a computer program to scan the reams of data, and obtaining an objective and quantitative answer to that one question while missing the more significant answers which are suggested only by entirely unexpected patterns in those endless details.[149]

In this quote, Richardson draws a gender division between "hypothesis-driven" science (done computationally) and a more intuitive and visual approach, reflecting the traditional gendering of experimentation as a male activity and natural history as female.[150]

Richardson and other protein taxonomists, in their various comparative approaches, experienced the same kinds of epistemic tensions as those who classified organisms in more typical natural history activities. From the 1930s to the present, proponents of different forms of "experimental taxonomy" have clashed with proponents of

[145] Richardson, "Beta-Sheet Topology and the Relatedness of Proteins," *Nature* 268 (1977): 495–500.
[146] Richardson, "The Anatomy and Taxonomy of Protein Structure," *Advances in Protein Chemistry* 34 (1981): 167–339.
[147] Ibid., 170.
[148] Cited in Bahar, "Ribbon Diagrams" (cit. n. 144), 5.
[149] Ibid., 6; emphasis in the original.
[150] Evelyn Fox Keller, *Reflections on Gender and Science* (New Haven, Conn., 1985).

morphological taxonomies over the issue of the objectivity of classifications.[151] Among the latter, Mayr, a leading systematist, valued subjectivity most, writing that the "good doctor and the good taxonomist make their diagnoses by a skillful evaluation of symptoms in the one case and of taxonomic characters in the other."[152] Simpson, a paleontologist, likewise argued that the identification of species depended "on the personal judgment of each practitioner of the art of classification." To this, he added that classification could not be objective: "To insist on an absolute objective criterion would be to deny the facts of life, especially the inescapable fact of evolution."[153] The experimental taxonomists, such as the molecular evolutionists, disagreed strongly with these assessments and insisted that classification and phylogeny should be objective and quantitative. They argued that reliance on molecular data, not morphology, was necessary to reach these goals.[154]

Similarly, in protein science, a number of researchers remained somewhat skeptical about the validity of visual methods to classify proteins. They developed alternative methods that they claimed would "analyze automatically and objectively" the coordinates of proteins to identify protein domains. These researchers criticized those who relied on the visual inspection of three-dimensional models stored in the PDB.[155] Automated approaches, they argued, also conducted with data from the PDB, were far superior because they were objective. Similar concerns with the objectivity of visual comparison were widespread.

These classifications of proteins, whether produced visually or automatically, borrowed (sometimes consciously, sometimes not) from standard natural historical practices. By the beginning of the twenty-first century, some protein scientists were ready to acknowledge the similarity between their work and that of naturalists. In a 2002 review titled "The Natural History of Protein Domains," protein researchers drew these parallels explicitly:

> For over a century, zoologists have classified organisms using the Linnaean system in order to provide insights into their natural history. Biologists are beginning to appreciate the benefits of hierarchical domain classification systems based on sequence, structure, and evolution. The numerous parallels between these systems suggest that domain classifications will prove to be key to our further understanding of the natural history of domain families.[156]

This is not to say that current practices in protein classification have returned biological research to its natural historical origins. Rather, collecting and comparing practices have been essential to both natural historical and experimental research. As seen in protein taxonomies, bringing modern databases into the larger framework of collections highlights problems shared equally by naturalist collectors and the database users, such as the place of subjectivity in comparing biological shapes, the role

[151] Strasser, "Laboratories, Museums" (cit. n. 95).
[152] Mayr, Linsley, and Usinger, *Methods and Principles* (cit. n. 75), 106–7.
[153] George Gaylord Simpson, *Principles of Animal Taxonomy* (New York, 1961).
[154] On this debate, see Strasser, "Laboratories, Museums" (cit. n. 95).
[155] M. Levitt and J. Greer, "Automatic Identification of Secondary Structure in Globular Proteins," *Journal of Molecular Biology* 114 (1977): 181–239.
[156] C. P. Ponting and R. R. Russell, "The Natural History of Protein Domains," *Annual Review of Biophysics and Biomolecular Structure* 31 (2002): 45–71.

of visual and numerical approaches, and the function of taxonomies in organizing collections.

In his recent *Styles of Knowing: A New History of Science from Ancient Times to the Present*, Kwa claims that "a comparative method is a means of building a taxonomy, nothing more, nothing less."[157] Actually, it is a bit less and a lot more. The production of taxonomies has perhaps been the most visible use of the comparative method as applied to collections, with examples ranging from Linnaeus's herbarium to Mayr's vertebrates at the American Museum of Natural History. But it has not been the only one. The comparative method has also served to identify specimens, relying on the fact that collections are embodiments of knowledge systems. Most obviously, type specimens (holotypes) stored in natural history museums serve as the ultimate referent for a species.[158] Naturalists compare specimens of unknown status to a type specimen in order to assess whether they belong to the same species or to another one. Modern databases have served an identical role. The most precious help that computerized databases of DNA sequences can provide to experimentalists and that a dispersed set of printed sequences cannot is in identifying the function, and thus the identity, of genes. DNA sequences are often determined before the function of a gene is known. In the case of a gene coding for a protein, instead of attempting to find experimentally every possible biochemical reaction in which that protein might be involved, researchers compare their new DNA sequences with all other sequences available in a database, using algorithms such as BLAST. And if they find a "match," that is, a sequence that is sufficiently similar, and whose function is known, they can infer that the two sequences produce proteins of similar structure and function.[159] This function can then be explored further experimentally. Databases offer a unique shortcut for experimental investigations: they suggest possible roles for proteins that scientists had never even thought of.[160] Although journals first accepted the evidence of sequence comparisons as sufficient to warrant a publication, they soon required that the results be confirmed experimentally. Yet sequence comparison remains a crucial step in the process of producing knowledge experimentally.

CONCLUSIONS

By the beginning of the twenty-first century, public collections of molecular data numbered in the thousands; so many, in fact, that databases of databases were established to help researchers keep track of these resources.[161] One significant question remains, at least from the perspective of a generalist vision: Although these twenty-first-century databases certainly resemble earlier natural history collections, do they bear any historical connections to them?

[157] Kwa, *Styles of Knowing* (cit. n. 24), 167.

[158] On the epistemology of type specimens, see Lorraine Daston, "Type Specimens and Scientific Memory," *Crit. Inq.* 31 (2004): 153–82.

[159] Michel Morange, *A History of Molecular Biology* (Cambridge, 2000), chap. 17.

[160] On the beginnings of this approach, see Russell F. Doolittle, "Some Reflections on the Early Days of Sequence Searching," *Journal of Molecular Medicine* 75 (1997): 239–41.

[161] M. Y. Galperin and G. R. Cochrane, "The 2011 *Nucleic Acids Research* Database Issue and the Online Molecular Biology Database Collection," *Nucleic Acids Research* 39, supplement (2011): D1–D6.

To take a metaphor from evolutionary theory, the databases that grew in the twentieth century are certainly *analogous* to earlier collections, because they share a similar structure and have a similar function.[162] Databases, like earlier natural history collections, are organized assemblages of standardized objects. The physical (or virtual) proximity of these objects, their mobility within the collection, the temporary order in which they are arranged, and the uniform format in which they are kept make it possible for the investigator to approach these objects comparatively. This comparative perspective is perhaps the most distinctive epistemic practice associated with all kinds of collections. It has been most important for natural history, especially studies of systematics and evolution. It has been much less relevant in most of the experimental approaches to life, which have relied on a different perspective centered on exemplary phenomena, usually produced in model organisms (more on this later). One interesting exception is comparative embryology in the nineteenth and early twentieth centuries, which was at the same time experimental and comparative (obviously). Comparative embryologists experimented on developing embryos, but (unlike physiologists, e.g.) they also relied extensively on collections of objects such as wax embryos, preserved tissues, and microscopic slides.[163] Here again, the collecting and comparative approaches were closely associated, in the same way as they have been associated around the experimental data and data collections examined in this article. The most striking similarity between modern databases and earlier collections in natural history is not simply that they are all collections of some sort, but that they have been constituted through similar collecting practices and have been put to use in similar ways for the production of knowledge.

One might object that databases and biological collections cannot be subsumed under the same heading because the former deal with information and the latter with material things. However, as noted above, it would be erroneous to equate natural history collections solely with the cataloging of whole or parts of organisms. Take Candolle's herbarium: it contained whole plants, although they were dried between sheets of paper, but for large plants, such as trees, it contained only parts of the plants, usually flowers and some leaves. It also contained, in the same format of large sheets, drawings of plants, or even verbal descriptions. This particular collection thus contained an entire ontological range of collected objects, from material things to abstract ideas. The difference between things and data is very real, but it is more a matter of degree than a matter of kind.[164]

Obviously, the contemporary biomedical sciences have not simply returned to the old methods of natural history. Current electronic databases differ in many ways from former natural history collections. But they have reincorporated collecting and comparing approaches into the experimental tradition. What is most distinctive about current biomedical research is its hybrid character that produces knowledge through both experimentation and collection.[165] Establishing this successful hybrid culture has not been simple, and incorporating collecting practices has had deep consequences for the entire research enterprise. It has forced a reexamination of notions of

[162] I thank Robert Kohler for suggesting this useful analogy.
[163] Nick Hopwood and Friedrich Ziegler, *Embryos in Wax: Models from the Ziegler Studio* (Cambridge, 2002).
[164] See Latour's wonderful essay on the ontological range of scientific objects in Latour, *Pandora's Hope—Essays on the Reality of Science Studies* (Cambridge, 1999), chap. 2.
[165] Strasser, "Laboratories, Museums" (cit. n. 95).

authorship, which can no longer be understood solely in individual terms as it was in the experimental sciences, and has brought about serious changes in attitudes toward data ownership and data sharing, challenging the very meaning of "publication."

If one accepts that these various collections are indeed analogous, there remains a question of whether modern databases are *homologous* to earlier natural history collections. In addition to sharing a common structure and function, do they share a common descent? Is there some historical connection, whether social, intellectual, or cultural, between collections and databases? This question is far more difficult to answer, although the most probable answer is *no*. Most data collectors of the twentieth century were not naturalists trained in the arts of collecting specimens in the field, preparing them for herbaria or museum collections, and using a broad, systematic comparative approach to produce knowledge; they were trained as experimentalists (or theoreticians of some sort). When they began collecting and comparing, they became alienated from their experimentalist colleagues, many of whom no longer considered them scientists. And naturalists did not have much more sympathy for these new collectors, many of whom had never been in the field.[166] The data collectors of the twentieth century went through the difficult process of creating a new professional identity for themselves. Only in the 1980s did science funding agencies fully recognize the importance of data collectors for the progress of the experimental sciences. At the same time, the term *bioinformatician* was gaining wider acceptance (the journal *Bioinformatics* was created in 1985); this term seems to point solely to the use of computers in biology, but in fact designates a professional group committed to producing knowledge through the collection and comparison of data.

If we accept, for heuristic purposes, that modern and ancient collections are at least analogous, though not homologous, how does this contribute to a generalist vision of the life sciences from the early modern period to the present? First, the rise of experimentalism did not mean an irreversible movement away from all other practices of biological investigation. Other ways of knowing (e.g., those based on collecting and comparing) that were centrally important in the early modern period continue to be central for naturalist research, but have also, as this article has argued, grown in importance in experimental research as of the beginning of the twenty-first century. The twentieth century can still be safely qualified as the "experimental century," but the twenty-first might well be a "collecting century." If one were to take an even broader view, the twentieth century might no longer be considered the culmination of methodological progress leading to experimentalism, but rather a brief, albeit significant, interlude in the history of the life sciences.

To be sure, this article does not have the pretension to replace the standard narrative crafted by Coleman and Allen with another narrative centered on collecting practices. Nor does it claim to have identified an actor's concept that historians have ignored. It claims that *collecting*, as an analytical category of practices, can be productively deployed in writing a history of the life sciences, not only in the field and the museum, where these practices have been most closely examined, but in the laboratory, where the focus has almost exclusively been on experimentation. It also suggests that this approach might be successful in connecting recent developments in the biomedical sciences, such as the proliferation of databases and data-driven

[166] See, e.g., field naturalists' attitudes toward the experimental taxonomist Alan Boyden, described ibid.

methods, diachronically with the earlier history of the life sciences and synchronically with sciences such as systematics or ecology, which have developed mainly outside the laboratory. Thus, instead of seeing these recent developments in data-driven science as yet another revolution (or worse, a "fourth paradigm") and isolating them from other changes taking place in other disciplines, one may succeed in producing a narrative that brings them into historical perspective and offers critical insights into these transformations.[167]

Seeing the persistence of collecting practices over time leads to another crucial question. Why did collecting approaches develop so prominently in the life sciences? Was it a historical accident, or does something about the objects studied by life scientists lend itself to comparative perspectives? One key reason why comparative perspectives and, thus, collections have been so important in the study of life is that natural selection operates on functions but is blind to structures. Evolution has resulted in a variety of structures performing similar functions, making it particularly difficult for researchers, whether the biological components they consider are molecules or morphological traits, to infer functions from single structures. Physicists and chemists do not have that problem: all the entities of one kind that they explore are believed to be structurally identical. Collecting and comparing, then, is intimately linked to the historicity of the objects investigated. This is borne out in considering the nature of other historical disciplines: geology and cosmology, like the life sciences, have both been heavily dependent on collecting practices.

Bringing collecting to the fore thus leads to new questions about the boundaries between the sciences. Rather than following historically contingent disciplinary boundaries, it might be more productive to think about the deep commonalities between the sciences that are historical (biology, geology, cosmology) and those that are not (physics, chemistry, mathematics). It can also help us question the divisions between the natural and the social sciences. What would happen if we began to think about the aims of biology and history on one side, and those of physics and sociology on the other? Writing history from the vantage point of specific practices, such as collecting, makes such connections visible, in a way that disciplinary histories do not.

Whether scholars follow up on this latter suggestion or not, I hope that this article has at least made clear that it can be productive to think about current databases as collections that follow a long tradition of collecting in the life sciences. Other readings are possible, of course, and worth exploring. One might, for example, bring databases into the context of the encyclopedist movement and the development of library sciences. But this article has offered a first attempt to ask some of the same kind of questions about modern databases that historians of natural history have asked about wonder cabinets, herbaria, and museums. This historical contextualization draws attention to the variety of collecting strategies, to the challenges of managing a heterogeneous network of collectors, and to the epistemic challenges of working comparatively, especially in an experimentalist age. The analytic focus on collecting practices (in the field *and* the laboratory) and the contrasts between experimental and relational systems, between exemplary and comparative perspectives, and between centrifugal and centripetal places make it possible to overcome the distinction between natural history and experimentation, the museum and the laboratory, and hypothesis-driven

[167] Bruno J. Strasser, "Data-Driven Sciences: From Wonder Cabinets to Electronic Databases," *Stud. Hist. Phil. Biol. Biomed. Sci.* 43 (2012): 85–7.

and data-driven science. It also helps us bring current claims about the uniqueness of contemporary science into perspective.

Indeed, broadening the perspective, one might question the role of the laboratory in defining modern science. In concluding *Reinventing Knowledge* (2008), a pointed overview of the six major institutions of knowledge developed during the last two and a half millennia, Ian F. McNeely and Lisa Wolverton claim that "by the mid-twentieth century, the laboratory had ascended to an almost impossibly dominant status as an institution of knowledge" and that "laboratory science and its accomplishments now act as the chief means by which Western knowledge systems manifest their superiority to the rest of the world" (Herbert Butterfield would have approved of the style and content).[168] The laboratory remains obviously indispensable to and powerful for the production of knowledge about the natural world, but the proliferation of data collections and comparative approaches seems to indicate that it no longer enjoys this dominant position alone. As McNeely and Wolverton show, the fortunes of different institutions of knowledge have changed over time. Museums, for example, which once "performed indispensable functions in legitimating knowledge," now thrive in different roles, such as "education, entertainment, and outreach."[169] But the epistemic qualities that made museums so vital to the production of knowledge—their role as stable referents of the natural world and the possibility of applying comparative approaches to their collections—are now also present elsewhere, in digital databases that might be thought of as "data museums."

Reframing the history of recent science in this perspective also illuminates the recent politics of knowledge. The increasing use of databases for the production of knowledge has only made the question of access more acute, a question that has been addressed extensively by naturalists in the case of natural history collections—for example, by defining rules about the borrowing of museum specimens.[170] Because databases were the product of broad collective efforts, many argued that they should be freely accessible and open to all. This position also facilitated the collection of data and maximized the potential use of databases. What represented a pragmatic decision for the managers of databases, such as GenBank, also lent support to broader initiatives to make knowledge more accessible. The success of GenBank's open-access policy was used as an argument to promote PubMed Central, an open repository of published scientific literature, and eventually the NIH's open-access policy (all publications resulting from NIH-funded research must be deposited on PubMed Central within a year).[171] The greater availability of scientific knowledge in a format that lends itself to the further production of knowledge has made possible a broader participation in science, including anyone from secondary school teachers and their students in the classroom to computer-game amateurs competing to solve protein-folding problems.[172] The availability of data collections and their increasing legitimacy for the production of knowledge has fueled the growth of "citizen science" and

[168] McNeely and Wolverton, *Reinventing Knowledge* (cit. n. 28), 251, 271; Butterfield, *The Origins of Modern Science, 1300–1800* (London, 1949).

[169] McNeely and Wolverton, *Reinventing Knowledge* (cit. n. 28), 256.

[170] On the norms about the borrowing of specimens, see Mayr, Linsley, and Usinger, *Methods and Principles* (cit. n. 75).

[171] R. J. Roberts et al., "Building a 'Genbank' of the Published Literature," *Science* 291 (2001): 2318–9.

[172] S. Cooper et al., "Predicting Protein Structures with a Multiplayer Online Game," *Nature* 466 (2010): 756–60.

made its expansion more plausible than ever before, especially for the experimental sciences.

As I have suggested in this article, looking at collections beyond their alleged decline in the late nineteenth century offers promising venues to contextualize some of the deep transformations currently taking place in science. For one, data-driven science now seems more familiar and less a product of our "information age." As Robert Darnton has reminded us, "Every age was an age of information, each in its own way,"[173] and the age when natural history was most flourishing was no exception. The extensive use of collections by naturalists and the existing scholarship on natural history collecting provides the historian of recent science with rich material to ask fresh questions about the use of current databases in science. The insights of earlier naturalists about the epistemic, social, and cultural challenges of working with collections help us understand some of the current difficulties faced by the participants in data-driven science. Indeed, when Simpson referred to "a science that is most explicitly and exclusively devoted to the ordering of complex data," he was not referring to current data-driven science but to animal taxonomy.[174] Same questions, different times.

[173] Darnton, "An Early Information Society: News and the Media in Eighteenth-Century Paris," *Amer. Hist. Rev.* 105 (2000): 1–35.
[174] Simpson, *Principles of Animal Taxonomy* (cit. n. 153), 10.