Journal of the History of Biology (2013) 46:401–444
DOI 10.1007/s10739-012-9336-6

Towards "A Natural History of Data": Evolving Practices and Epistemologies of Data in Paleontology, 1800–2000

DAVID SEPKOSKI
Max Planck Institute for the History of Science
14195 Berlin
Germany
E-mail: dsepkoski@mpiwg-berlin.mpg.de

Abstract. The fossil record is paleontology's great resource, telling us virtually everything we know about the past history of life. This record, which has been accumulating since the beginning of paleontology as a professional discipline in the early nineteenth century, is a collection of objects. The fossil record exists literally, in the specimen drawers where fossils are kept, and figuratively, in the illustrations and records of fossils compiled in paleontological atlases and compendia. However, as has become increasingly clear since the later twentieth century, the fossil record is also a record of data. Paleontologists now routinely abstract information from the physical fossil record to construct databases that serve as the basis for quantitative analysis of patterns in the history of life. What is the significance of this distinction? While it is often assumed that the orientation towards treating the fossil record as a record of data is an innovation of the computer age, it turns out that nineteenth century paleontology was substantially "data driven." This paper traces the evolution of data practices and analyses in paleontology, primarily through examination of the compendia in which the fossil record has been recorded over the past 200 years. I argue that the transition towards conceptualizing the fossil record as a record of data began long before the emergence of the technologies associated with modern databases (such as digital computers and modern statistical methods). I will also argue that this history reveals how new forms of visual representation were associated with the transition from seeing the fossil record as a record of objects to one of data or information, which allowed paleontologists to make new visual arguments about their data. While these practices and techniques have become increasingly sophisticated in recent decades, I will show that their basic methodology was in place over a century ago, and that, in a sense, paleontology has always been a "data driven" science.

Keywords: Fossil record, Databases, Paleontology, H.G. Bronn

Introduction: The Fossil Record and the Archive

One of the major transformations in the history of paleontology occurred during the second half of the twentieth century, when the new sub-discipline of "paleobiology" emerged, centered around the quantitative analysis and interpretation of the history of life (Sepkoski, 2012). A central feature of the emergence of paleobiology was the increasing importance of large quantities of data as a source of this analysis, and, since the late 1970s, computer databases have become essential tools in paleontology. An exemplar of this phenomenon is the collaborative "Paleobiology Database," a project established in 1998 by researchers at a number of different institutions around the world. Essentially, the Paleobiology Database is a massive, electronic clearinghouse of taxonomic and stratigraphic information about the known fossil record that is freely accessible via the internet. The project website describes the database as

> a public resource for the global scientific community. It has been organized and operated by a multi-disciplinary, multi-institutional, international group of paleobiological researchers. Its purpose is to provide global, collection-based occurrence and taxonomic data for marine and terrestrial animals and plants of any geological age, as well as web-based software for statistical analysis of the data. The project's wider, long-term goal is to encourage collaborative efforts to answer large-scale paleobiological questions by developing a useful database infrastructure and bringing together large data sets.[1]

This collaboration was a natural evolution of paleontological database projects that began in the 1970s and 1980s, which had the goal of gathering the results of some 200 years of fossil collecting and describing into a single database that could claim to represent, in idealized form, the complete "fossil record." The most famous such database is the one compiled by J. John Sepkoski, Jr. for the Phanerozoic marine fossil record between roughly 1975 and 2000, which compiled data for marine fossils at taxonomic level of the family and,

[1] http://paleodb.org/cgi-bin/bridge.pl?a=displayPage&page=paleodbFAQ, accessed 30 November, 2011.

eventually, the genus (Sepkoski, 1982, 1992; Sepkoski et al., 2002).[2] As digital storage has become cheaper and information more widely accessible via the internet, fossil databases such as Sepkoski's have grown enormously in size: the Paleobiology Database contains some 240,000 taxa culled from over 130,000 collections and 42,000 bibliographic references, representing over a million taxonomic occurrences. Database projects such as these also foster a degree of "collective empiricism" unparalleled in the history of paleontology, as researchers can access and contribute to these databases from literally anywhere in the world (Daston and Galison, 2007). In this sense, modern paleontology seems emblematic of what some observers have argued is the emergence of "data-driven" science: an ostensibly recent transformation in which scientific practice has become increasingly centered around massive sets of data and dependent on technologies that facilitate management and analysis of that data.[3]

From its beginnings as a professional discipline, paleontology has been dependent on its collections, and it remains so today. Since the early nineteenth century, paleontologists have accumulated large numbers of objects – fossils – that illuminate the history, function, morphology, and environments of extinct organisms. Then, as now, fossils have been collected, cleaned, and identified; specimens have been tagged and stored in collection drawers for reference and retrieval; descriptive and taxonomic monographs have been published to aid research and pedagogy. This activity has produced the central resource of paleontology – the "fossil record" – from which paleontologists have abstracted their knowledge of the patterns and processes of evolution and extinction in the history of life. However, if you were to ask a geologist in 1830 or 1840 to see the fossil record of life, you might be shown to a row of specimen drawers or, more likely, handed a lavishly illustrated book featuring engravings of typical specimens from various strata or regions, often accompanied by a list of taxonomic groups arranged in Linnaean systematic order. If you asked the same thing of a paleontologist in 2012, you would probably be shown to a computer. There you could access a database consisting of tens of thousands of entries

[2] In the interests of full disclosure, J. John "Jack" Sepkoski, Jr. (1948–1999) was the author's father. For a reflection about how I approach writing history about a close family member, see Sepkoski, 2012, pp. 6–7.

[3] While it is difficult to precisely define the exact characteristics of data-driven science, two of its central features are, as Sabina Leonelli has noted, a methodology that prioritizes induction from accumulated data (rather than deduction from preconceived hypotheses), and automated information processing and analysis, usually facilitated by computers (Leonelli, 2012, p. 1).

compiled from a global survey of all taxonomic literature produced over the last two centuries that could be sorted, subjected to statistical analysis, and from which graphical images of patterns of evolution, diversification, and extinction over time could be generated. Both examples could be described as representations of "the fossil record."

What this comparison shows is that while the basic collection practices of the discipline have not changed in almost 200 years, what is *done* with those collections – the analytical practices, the epistemic values, the representations of knowledge, and the technologies and media that support the process – have evolved significantly. By the later twentieth century, information about fossils has come to be conceived as a record of pure data, in which physical fossil specimens have been reduced to interchangeable data points that can be digitized, analyzed statistically, and used as the basis for mathematical models and simulations. Since the 1970s these databases have become ever more massive, and computers have become a central tool in paleontological data analysis, serving as the basis for what is sometimes described as a "paleobiological revolution" (Sepkoski and Ruse, 2009).

This paper examines the relationship between the changing material forms that knowledge of the fossil record has taken (e.g. cabinet, book, database), and the analytical practices and epistemic values associated with them. I will trace how the "fossil record" has functioned both as a collection of things (fossils) located in actual or virtual physical space, and as a database of information, a collection of data points not tied to specific physical objects. One task is to historicize the emergence of the database by examining the history of data practices in paleontology, focusing especially on a moment in the mid-nineteenth century when new quantitative and visual strategies emerged to cope with increasing quantities of paleontological data that were being collected and analyzed. I will argue that, at least in the case of paleontology, while knowledge has always been derived from collections, the nature of these collections and their meanings has changed over time. These collections contained, successively, objects (fossils), illustrations, numerical tables, graphs, and electronic data, and at each point can be considered both specific representations of nature and at the same time a "second nature" from which knowledge is produced (Latour, 1999, Chap. 2). However, one of the central epistemic values associated with modern paleontological databases – quantification and numerical analysis – emerged long before the era of computer technology and advanced statistical techniques. In this sense, I will join those who challenge the notion that data-driven science is a wholly recent development, or that its arrival was entirely dependent on modern technologies (Strasser, 2012; Müller-Wille and Charmantier, 2012).

I will also show that the emergence of certain kinds of visual (though often non-pictorial) representations was embedded in this shift. These non-pictorial representations (such as diversity curves and "spindle diagrams") were tools that allowed paleontologists to make new visual arguments about their data, and they helped introduce a visual hermeneutics for analyzing the massive amounts of data that were beginning to accumulate in the fossil record. The urge to catalog objects of nature is common to all branches of natural history as part of a historical trend that extends back to antiquity.[4] However, by the early nineteenth century, when paleontology emerged as a professional discipline, new forms for representing data emerged that reflected new disciplinary and epistemic concerns. What was initially a pictorial or descriptive approach to cataloging individual fossils gradually gave way, as paleontological knowledge expanded, to quantitative and non-pictorial representations of the fossil record that served new uses to which fossil data were put. Compilations of the fossil record began as sites for depicting a typology of objects in stratigraphic or physical space, but ultimately became compendia of taxonomic entries in which objects were reduced to data points that could be stored, retrieved, and manipulated. In this way, the collection catalog evolved to become the modern digital database. What is particularly surprising about this history, however, is that the beginnings of this transformation did not depend on the arrival of computers in the later twentieth century. Rather, distinct interpretations of the fossil record as both a record of data and as an collection of objects have coexisted for most of the history of the professional discipline of paleontology.

The Fossil Record as Catalog

The modern digital databases used today bear little obvious relationship to the earliest fossil compendia, which date from the late sixteenth century explosion of Renaissance natural historical texts. In the first place, during that early period the term "fossil" referred to all kinds of natural objects found in the earth, including minerals, crystals, and

[4] As a number of authors have pointed out, Early Modern naturalists had their own "data deluge" to cope with, and developed strategies and technologies that were often quite sophisticated. See Ogilvie, 2003; Blair, 2010; Müller-Wille and Charmantier, 2012.

strangely-shaped stones, as well as what we now recognize as the petrified remains of organisms. It would be anachronistic, therefore, to refer to such early works as 'paleontological,' or to call them representations of "the fossil record," since naturalists of this era neither recognized the distinction between organic and non-organic fossils, nor had a conception of deep historical time associated with the emergence of geology in the eighteenth century. Nonetheless, between roughly 1550 and 1600 some of the earliest descriptions and illustrations of fossil specimens appeared in texts published by naturalists including Georg Bauer (1494–1555, known more commonly as Agricola), Conrad Gesner (1516–1565), Ulysse Aldrovandi (1522–1605), Jean Bauhin (1541–1613), and others.

The history of early works on fossils has been ably discussed by a number of historians, most prominently Martin Rudwick, and I will not linger on it here (Rudwick, 1976; Rappaport, 1997). However, several important features of early fossil collection and description practices emerged in this period that bear on the later history of fossil compendia. The first and, as Rudwick notes, most important innovation was the marriage of images to text, which allowed for more precise naming and identification of fossils than mere verbal description, and also emphasized the importance of first-hand observation by the naturalist of the objects being described (Rudwick, 1976, pp. 5–10; 2000). This innovation was closely related to developments in print technology that allowed more and more sophisticated images of fossil objects to be mechanically reproduced, beginning with the comparatively crude woodcut illustrations of the sixteenth century, and progressing through the introduction of copper engraving, etching, lithography, and eventually photography in the late nineteenth century.[5] A second important feature was that later sixteenth-century works on fossils were increasingly based on large physical collections of specimens, and the works themselves served as proxies for those physical collections. As Rudwick points out, published illustrations of fossils were "a convenient substitute for a museum," and sixteenth century illustrated works on fossils were contemporary with the establishment of the first important collections of fossils by private individuals. Illustrated works allowed the wide duplication and dissemination of these collections, "thereby placing the same data at the disposal of naturalists everywhere." Rudwick notes that this was especially important in paleontology, since "even the commonest fossils generally have to be collected from

[5] On the history of refinements in natural history illustration techniques, see Rudwick, 1975, 1976; Davidson, 2008; Daston and Galison, 2007.

extremely restricted localities… which may not be known or accessible to any but those living close by'' (Rudwick, 1976, p. 11). This feature is closely related to the third innovation, which is that published fossil compendia helped promote the study of fossils as a collective enterprise, and encouraged cooperative research. Beginning with the earliest of these works, such as Gesner's *De rerum fossilium* (1565), authors encouraged the active sharing of information on fossils from regions or localities that would be difficult for most naturalists to physically visit (Rudwick, 1976, p. 14). By the nineteenth century, fossil collections based on regional sampling were widely dispersed around the cosmopolitan research centers and university towns of Europe, and published atlases of these individual collections served as the glue for the kind of ''collective empiricism'' described by Lorraine Daston and Peter Galison that unites disciplines around ''common objects of inquiry'' (Daston and Galison, 2007, p. 22).

What the earliest works on fossils did not typically convey, however, was a sense of fossils as belonging to natural groups, or of being part of a historical progression. By the eighteenth century, naturalists began to make more systematic attempts to catalog large collections of fossils, and to place the fossil organisms in some kind of taxonomic order. Eventually, these fossil catalogs became tools that supported the great project of establishing the true order of the earth's strata and fixing the geological timescale. Stratigraphy emerged as a science in the late eighteenth century, when geologists like Abraham Werner, William Smith, and Georges Cuvier began to realize that fossils could be used as markers for determining the order of strata in the earth (Rudwick, 2005). Catalogs that identified and classified fossils according to taxonomic and stratigraphical position exploded in the first half of the nineteenth century, where they served as important reference tools while geologists attempted to produce a universal stratigraphy for Europe and the rest of the world. This would introduce an important convention: whereas earlier works had been concerned with determining the nature of the fossils themselves, in the era of stratigraphy the interest in fossils was secondary to the problem of determining the order and timescale of the geologic record.

One of the first pre-stratigraphic attempts at a systematic catalog of organic fossils in a particular region was the English divine John Woodward's *Towards a Natural History of the Fossils in England* (Woodward, 1729). This work was the fruition of decades Woodward spent collecting fossils of all kinds (of both organic and inorganic origin), and the collection itself was bequeathed to the University of

Cambridge, where it became the nucleus for the eventual Sedgwick Museum of geology. In his catalog, Woodward took pains to precisely identify the circumstances in which his fossils were found, noting ‘‘the Condition of the Earth, and of the Strata, not only in the places where these Bodies were found, but all others, where the Intrails of the Earth happened to be, by any means, display'd and laid open to view'' (Woodward, 1729, Part II, p. 2). Woodward's work contributed to a growing recognition that certain fossils were found only in particular strata, a realization would make the science of stratigraphy possible. He also presented his catalog of fossils in a taxonomic arrangement, noting the ‘‘class'' and ‘‘species'' to which each fossil belonged (although these categories should not be confused with later, Linnaean classifications). Woodward's epistemic concerns were very different than those of later paleontologists, who classified fossils in order to document the historical progression of geological strata, and eventually to reconstruct the historical record of life. But Woodward's work helped establish standard conventions for fossil compendia; it was, as the publisher's preface to the volume somewhat grandly claimed, ‘‘a Work exceedingly wanted in the world, yet scarce ever attempted… as enabled the *Doctor* to methodize them [fossils] according to their several *Species*, and reduce them into a Science'' (Woodward, 1729, pp. v–vii).

If, as Sandra Herbert puts it, before the 1830s ‘‘fossils were attended to primarily for their value as markers of strata, rather than for their purely biological significance,'' stratigraphy nonetheless introduced an important element of historicity to interpretations of the fossil record (Herbert, 2005, p. 81). Along with earlier contributions to eighteenth century geology by Georges Louis Leclerc (Comte de Buffon), James Hutton, and others, the development of stratigraphy coincided with what historians have referred to as ‘‘the discovery of deep time,'' or the realization that (a) the earth is very old, (b), its history is directional and shaped by geological processes that can be understood outside the context of sacred or human history (Rudwick, 2005). Initially, the fossils used as markers for stratigraphy were cataloged without any genealogical preconceptions; however, once the time-stratigraphic system was in place, it was a simple matter to connect the representative fossils as being part of a historical progression. This enabled a further epistemic shift in paleontology: the interpretation of fossils as being part of a biological succession now became central, because life was understood as having had a long history, extending back through successive stages of the history of the earth. It was, in a sense, at this point that the notion of a ‘‘fossil record'' first emerged.

Many of the fossil catalogs produced during the first several decades of the nineteenth century as part of the stratigraphic project convey this emerging sense of the historical and nature of the earth. William Smith, the pioneer of stratigraphy who produced the first geological maps of England, clearly understood that an analogy could be made between the organization of fossils in cabinets and catalogs and the original manner in which they were deposited in the earth's crust: "The organized fossils… may be understood by all, even the most illiterate, for they are so fixed in the earth as not to be mistaken or misplaced; and may be as readily referred to in any part of the course of the Stratum which contains them, as in the cabinets of the curious" (Smith, 1816, p. 1). Scottish geologist Robert Jameson explained, in his 1813 preface to his edition of Cuvier's *Essay on the Theory of the Earth*, that "Petrifactions [fossils] are no longer viewed as objects of mere curiosity, as things isolated and unrelated to the rocks of which the crust of the earth is composed; on the contrary, they are now considered as one of the most important features in the strata of all regions of the earth." This was, he argued, primarily because they demonstrate a "gradual succession in the formation of animals… and makes [the naturalist] acquainted with a geographical and physical distribution of organisms very different from what is observed to hold in the present state of the organic world" (Cuvier and Jameson, 1813, pp. vii–viii).

Geologists observed a variety of conventions when composing their fossil catalogs during this period, but two features in particular stand out as the most important and stable. First, a form of textual identification of fossil specimens became standard that generally included taxonomic identification (class, family, genus, and species), along with information about the stratum and locality where the specimen was found, and a description of particular morphological features that might aid in identifying the fossil. Eventually, as fossil compendia become less regional and more global, bibliographic references to other published descriptions of the fossils were also provided. Second, a standard form for visually depicting the fossils emerged, where representatives of a particular fossil type were grouped on a single, full-page plate, often with numbers keying the individual specimens to the textual descriptions in the volume (Figure 1). The composition of these images is noteworthy: the fossils are nearly always depicted top-down, neatly arranged in groups of similar genera or species, and framed within a rectangular border. This is precisely how fossils are arranged in museum cabinets, and these illustrations are, in effect, idealizations of the specimen drawers in which the collections being described were actually arranged. This format put the viewer in the advantageous position of being able to view the specimens just as if he

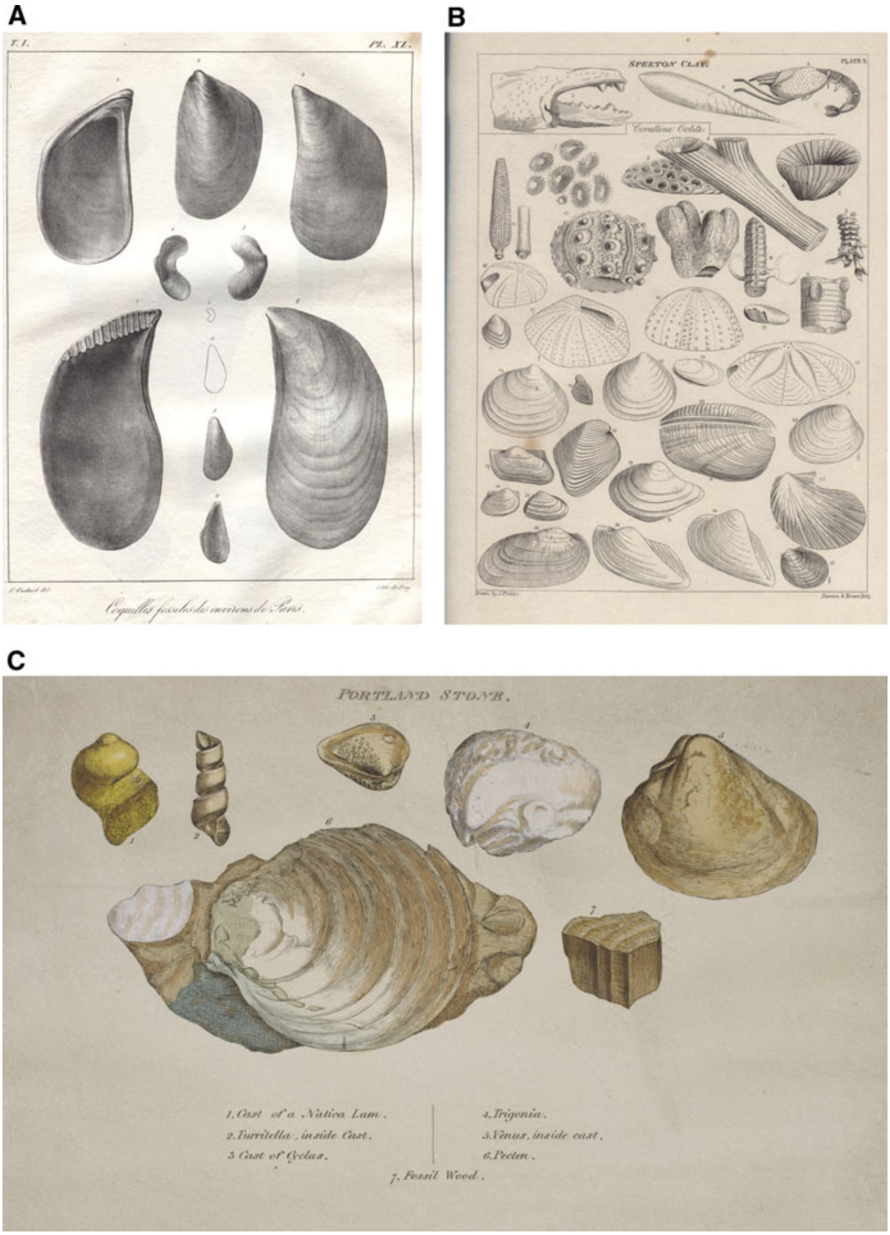

Figure 1. Early nineteenth century depictions of fossils in virtual ''specimen drawers.'' (**A**) Deshayes, 1824, Plate XL. (**B**) Phillips, 1835, Plate 3. (**C**) Smith, 1816, no page – bet. 16–17 (Image courtesy of the University of Chicago Library Special Collections Research Center)

were present to the physical collection; it also shows that, in many ways, the catalogs themselves were becoming – especially with the advent of more global compendia derived from multiple individual collections – virtual representations of the fossil record.

The Fossil Record as Table

During the 1820s and the 1830s, a large number of such elaborately illustrated fossil atlases were produced. In addition to Smith's and his nephew John Phillips' catalogs of fossils found in different regions of England, important surveys included Gideon Mantell's *The Fossils of the South Downs* (1822), James Parkinson's *Outlines of Oryctology* (1822), G.P. Deshayes' *Description des Coquilles Fossiles des Environs de Paris* (1824), and Adolphe Brongniart's *Histoire des Végétaux Fossiles* (1828) (Mantell, 1822; Parkinson, 1822; Deshayes, 1824; Brongniart, 1828a). However, beginning in the 1830s, a new genre of fossil compendia began to appear, with the publication of a number of works that consisted almost entirely of taxonomic lists of specimens, together with their stratigraphic location, without any illustrations at all. This is not to say that this period marked the end of illustrated fossil atlases; on the contrary, from the 1840s onward, ever larger illustrated compendia were produced, such as Alcide D'Orbigny's massive *Paléontologie Française*, a multi-volume publication that included hundreds of pages of plates depicting fossil specimens, and Deshayes' *Description Animaux sans Vertèbres*, which was similar in size and scope (d'Orbigny, 1840; Deshayes, 1860). But the emergence of a new genre of compilation that eschewed visual depiction of fossils marks an important transition in paleontology.

A prime early example of this new mode of catalog was Samuel Woodward's *A Synoptical Table of British Organic Remains* (1830), a modest text consisting of fewer than 100 pages in which fossils were arranged by taxonomic category and listed with information about their locality, stratum, and bibliographic reference in a tabular form (Figure 2). In his preface, Woodward (no relation to John Woodward) stated his hope that he was "rendering a service to the science of Geology," since no such general survey of British fossils had been made since John Woodward's 1729 *Attempt towards a Natural History of the Fossils of England* (Woodward, 1830, p. vii). One of the chief features of Woodward's *Synoptical Table* is that it was not based on first-hand collection experience, but was rather collated from the published catalogs and works of other English geologists. Just over a decade later, John Morris published a similar work entitled *A Catalogue of British*

MOLLUSCA. 23

CLASS, *Family*, Genus and Species.	Reference.	Strata.	Localities.
Spirifer obtusus	Min. Con. t. 269	Mt. Limestone	Scalebar, Yorkshire
lineatus	.. t. 493	..	Dudley, Worcester.
cuspidatus β	.. t. 461	..	Bristol. [Derbysh.
oblatus	.. t. 268	..	Axton Quarry, Flintsh. &
pinguis	.. t. 271	..	Black Rock, Ireland.
attenuatus	.. t. 493	..	Black Rock, Dublin.
bisulcatus	.. t. 494	..	Ibid.
distans	.. ..	..	Ibid.
rotundatus	.. t. 461	..	Torquay; Limerick.
Urii	Fleming, p. 376	..	Rutherglen, Lanark.
exaratus	.. ..	..	West Lothian.
decurrens β	G. T. 2 ser. iii. p. 164	Mag. Limestone	Torquay, Devon.
multiplicatus	.. p. 119	..	Humbleton, Tunstall Hill.
minutus	.. ..	..	Ibid.
Walcottii	Min. Con. t. 377	Lias	Bath; Hebrides.
Pentamerus Knightii	.. t. 28	Trans. Limestone	Downton, Wiltshire.
lævis	.. ..	Mt. Limestone	Bildwas, Salop.
Aylesfordii	.. t. 29	..	Colebrook D.; Croft [Ambery.
MOLLUSCA GASTEROPODA			
Phyllidiana			
Chiton octovalvis?	M. S. Nor. Crag	Crag	Thorpe, Norwich.
Patella lævis	Min. Con. t. 139	Lias	Whitby.
rugosa	.. ..	Inferior Oolite	Minchin-Hampton.
latissima	.. ..	Coal of Oolite	Bolingbroke, Linc.
ancyloides	.. t. 484	Great Oolite	Ancliff, Wiltshire.
Nanus	.. ..	..	Ibid.
lata	.. ..	Forest Marble	Stonesfield, Oxon.
locris	Geol. Eng. p. 129	L. or U. Gn. Sand?	Blackdown.
striata	Min. Con. t. 389	London Clay	Stubbington, Hants.
æqualis	.. t. 139	Crag	Holywells, nr. Ipsw.
unguis	.. ..	..	Ibid.
ferruginea, jun.	M. S. Nor. Crag	..	Bramerton, Norf.
Calyptraciana			
Emarginula tricarinata	Min. Con. t. 519	Great Oolite	Ancliff, Wilts.
scalaris	.. ..	..	Ibid.
crassa	.. t. 33*	Crag	Ipswich.
reticulata	.. ..	..	Holywells, nr. ditto.
Fissurella clathrata	.. t. 519	Great Oolite	Ancliff.
græca?	.. t. 483	Crag	Ipswich.
Pileopsis vetusta	.. t. 607	Mt. Limestone	Queen's Co. & Pres-
tubifera	.. ..	..	Ibid. [ton.
Pileolus plicatus	.. t. 432	Great Oolite	Ancliff; Hinton.
lævis	.. ..	..	Ibid.
Calyptræa trochiformis	Geol. T. ii. p. 204	London Clay	Hordwell Cliff.
Sinensis	Org. Rem. iii. t.5, f.10	Crag	Harwich, Essex.
Ancylus elegans	Min. Con. t. 533	Fresh Water	Hordwell Cliff.
Infundibulum echinulatum	.. t. 97	Plastic Clay	Plumstead, Kent.
obliquum	.. ..	London Clay	Barton Cliff, Hants.
tuberculatum	.. ..	..	Ibid.
spinulosum	.. ..	..	Ibid.
rectum	.. ..	Crag	Holywells, nr. Ipsw.
tenerum	M. S. Nor. Crag	..	Postwick, Norfolk.
Bullæana			
Bulla elongata	Phillips, t. 4, f. 7	Coral Rag	Scarborough.

Figure 2. Samuel Woodward's tabular arrangement of fossils (Woodward, 1830, p. 23)

Fossils (1843), which attempted to expand upon Woodward's *Synoptical Table* with a broader survey of published literature; this work was updated and enlarged in 1854 (Morris, 1843, 1854). Similar works were published during this time on the continent as well, including Brongniart's *Prodrome d'un Histoire des Végétaux Fossiles* (1828), Heinrich Bronn's *Italiens Tertiär-Gebilde und deren Organische Einschlüsse* (1831), and G.G. Giebel's *Fauna der Vorwelt* (1847) (Brongniart, 1828b; Bronn, 1831; Giebel, 1847). In some cases – such as Brongniart's and Giebel's works – the taxonomic lists were accompanied by a significant amount of descriptive text, but all of these compilations were broad surveys, collated from other published literature, that shared a basic commitment to providing systematic and stratigraphic data about fossils with little or no visual illustration.

What motivated the publication of these synoptical catalogs, and what did they reveal about the changing epistemic concerns of paleontologists? Initially, the earlier English works appear to have been motivated by the desire to aid in the project of general stratigraphy, essentially by providing an index of fossils found in particular strata. However, there are indications that their authors were beginning to envisage their compilations as useful for a broader study of the distribution and diversity of ancient life. As Morris put it in the preface to the first edition of his *Catalogue of British Fossils*, "the strict determination of species peculiar to each formation" would help "to enlarge the knowledge of the geographical and geological distribution of British Fossil Remains" (Morris, 1843, p. iii). In the preface to the second edition, he quoted the British geologist (and President of the Geological Society of London) Edgar Forbes' (1815–1854) comment that

> The value of palaeontology to the geologist depends on the evidence it affords of the continuity of species in time, which is the evidence of unbroken sequence of conditions; and the continuity of the group, which is the evidence of sequence of design. The names of species and genera are the words of the language by means of which such general facts and laws are expressed. (Forbes, in Morris, 1854, p. iii)

This indicates an inversion of an earlier logic, where fossils were interesting primarily for what they could reveal about geological history, to one in which stratigraphic information was useful for what it could reveal about the history of past life (Herbert, 2005, p. 86).

An epistemic holdover from the development of stratigraphy was the growing notion that geology – or paleontology, as by the 1830s the discipline was now being called – should involve the search for general laws and principles. In his *History of the Inductive Sciences* (1837), William Whewell spent over a hundred pages discussing the emergence of the ''palaetiological'' sciences, which included geology and paleontology, and whose object was to ''ascend from the present state of things to a more ancient condition, from which the present is derived by intelligible causes'' (Whewell, 1837, p. 481). Whewell's methodological prescriptions agreed with other contemporary philosophical accounts of the goals of science, and most prominently with developing notions of the proper *Wissenschaftliche* approach to natural history that were being promoted on the Continent, particularly by Alexander von Humboldt and his followers (Browne, 1983, p. 58; Gliboff, 2008, pp. 36–37). It was not unusual for scientists of this period to speak of science as the search for laws of nature; what was distinctive at this time was the emerging sense that branches of natural history – with their messy and heterogeneous sources of data – could aspire to producing such laws.

A number of fossil compendia of the 1830s and 1840s refer to establishing paleontology on a more secure epistemological basis by reforming and adding rigor to methodological practices. For instance, in his *Paléontologie Française*, D'Orbigny criticized earlier compilations' ''casual and incomplete figures and citations'' for producing uncertainty about the ''true limits of species'' and their distribution in the rocks. He proposed that his own heavily-illustrated atlas, which provided ''a strict review of the same species, exact figures, and accurate citations will eliminate any uncertainty, and bring the science to a positive and logical basis'' (D'Orbigny, 1840, p. 9).[6] But some paleontologists wanted to go further than just more accurately describing and illustrating fossil specimens; they viewed fossils as data supporting a general and even potentially quantitative understanding of the history of life. As a result of the process of compiling ever more global compendia of information about the stratigraphical locations, geographical distribution, and taxonomic order of fossil taxa – initially for the purpose of establishing the time-stratigraphic model – some paleontologists began to realize that they were accumulating something akin to a database of information about fossils that could be subjected to quantitative analysis of patterns of relative duration and diversity of organisms in the history of life.[7]

Rudwick has described what he has called British geologist Charles Lyell's ''dream of a statistical palaeontology,'' which he interprets as a

[6] All translations are mine unless otherwise noted.

[7] While, strictly speaking, the term ''database'' did not emerge until well into the twentieth century (the *Oxford English Dictionary* cites its first appearance as 1962), I use the term here for heuristic purposes to describe any large collection of abstracted data.

"failed" theoretical project undertaken in the late 1820s to produce a quantitative, general "faunal chronometer for the whole of the fossil record" (Rudwick, 1978, p. 236). The essence of Lyell's approach was to analyze the percentage of living versus extinct species of mollusks in all of the Tertiary basins of Europe, in order to produce a relative geological timescale based on the average longevities of species. Lyell collaborated on this project with the French mollusk expert Paul Deshayes, and the two compiled a catalog of over 3,000 Tertiary fossils as the basis for their analysis. Lyell hoped that by establishing the relationship of extant species in any formation to the age of the formation, a more general model of geological chronology could be established. Ultimately, Rudwick concludes that this project stalled in part because of Lyell's simplistic statistical assumptions and because of a variety of empirical difficulties in defining species limits, and he notes that the agenda failed to catch on with contemporary paleontologists (Rudwick, 1978, p. 241).

However, Lyell did mention his statistical project in the 1833 third volume of his *Principles of Paleontology*, where he compared the task of the paleontologist to a government census-taker, drawing an analogy between "the mortality of the population of a large country" and "the successive extinction of species, and the births of new individuals [to] the introduction of new species" (Lyell, 1830–1833, vol. 3, p. 31). Geoffrey Bowker regards Lyell's approach here as a crucial development in the history of epistemologies of data and archival practices in the sciences, arguing that Lyell's fossil 'census' "drew on the analogy of the information practices of statistics developed in large-scale government in the late eighteenth and early nineteenth centuries," and effectively served as "a kind of bookkeeping device that allows the storage of vast amounts of information by sorting them into a kind of filing cabinet of different kinds of event" (Bowker, 2005, pp. 55 and 67). This involved, Bowker maintains, "explicitly developing the concept that the earth formed its own archive," or "information storage device," and the belief that statistical techniques could be applied to this archive in order to "demonstrate underlying lawlike regularities in the face of current empirical chaos" (Bowker, 2005, pp. 55–57).

While Bowker's interpretation of this transformation may be overly broad, I am inclined to agree that Lyell participated in an important methodological innovation in paleontological data practices that was taking place at the time. Lyell may have abandoned his "statistical dream" for paleontology, but quantitative analysis of fossil data was an area of growing interest in paleontology between the 1830s and the 1860s. During this time, a number of works appeared that went beyond merely listing fossil taxa by stratigraphic location and systematic classification, but rather attempted genuine numerical meta-analysis of the data itself. This often came in the form of tables listing, for example, the number of species or genera for a particular group associated with a given geological period, from which some attempt was made to reach conclusions about species longevity, geographic distribution, and even processes of competitive replacement and extinction.

The basic methodology involved in these compilations was very similar to what Janet Browne has described as "botanical arithmetic," a term associated with Humboldtian studies of biogeography popular in the 1820s and 1830s among botanists and zoologists in Germany and France. As Browne explains, botanical arithmetic was "an elementary numerical technique that reduced absolute figures into statements of a proportional kind, which could then be arranged with others in a table" (Browne, 1983, p. 59). A common example of this practice was to calculate the ratio of one species to another, in order to identify the predominant species in an area, to illuminate relationships between species, or to indicate the relative incidence of particular taxa. Browne argues that this method became dominant in biogeographical studies after 1830, and that it eventually made inroads in paleontology and zoology. She also connects it closely with Humboldt's ideas about true *Wissenschaft*, since "to become numerical was essential, in his opinion, if the study of distribution was to be based on indisputable facts" (Browne, 1983, p. 60). However, Browne, like Rudwick, views these early "statistical" forays as ultimately futile, and reports that their popularity had waned by the 1850s, when such quantitative approaches were often derisively referred to as "Tabellenstatistik." She observes that, in most cases, "the purpose of the exercise rested in the figures, not in the conclusions which might be drawn from them," and argues that "to modern eyes many of these numerical surveys seem somewhat pointless," since they "were rarely used to substantiate specific hypotheses, nor did they generate any important new questions about geographical phenomena" (Browne, 1983, pp. 73 and 80).

Bronn and a Fossil Record of Data

Botanical arithmetic seems to have come to paleontology initially through Brongniart's great survey of fossil plants, *Histoire des végétaux fossiles* (1828), and the analytical *Prodrome* that accompanied it, where Brongniart engaged in some elementary numerical tabulations of fossil classes. The method was continued by Alphonse de Candolle, who refined some of Brongniart's studies of the ratios of plant species in different geological eras; by Deshayes, who assisted Lyell with his own statistical survey; and by John Phillips, who applied basic numerical calculations to his regional survey of British Paleozoic fossils (Phillips, 1841). But one of the most thorough converts to the numerical approach was the German paleontologist H.G. Bronn who, I argue, pushed the approach well beyond mere ''Tabellenstatistik'' into a genuinely new and important way of conceptualizing the fossil record.

Bronn's first major foray into numerical study of fossils was his 1831 study of Italian Tertiary fossils, *Italiens Tertiär-Gebilde*, which combined a lengthy non-pictorial catalog of fossil taxa with a number of elaborate numerical tables that demonstrated the quantitative relationships between the groups studied. In the introduction, Bronn stated his intention to ''give the most complete overview of the fossil remains of sub-Apennine formations, both of those which I collected myself, or had the opportunity to see, as well as those which I know only from reliable authors'' (Bronn, 1831, p. 1). His taxonomic catalog, then – which runs to some 140 pages – was compiled in part from a monographic survey of works by nearly 40 authors, including surveys by Basterot, Brongniart, Cuvier, Giovanni Brocchi, Deshayes, Lamarck, and D'Orbigny; in future works, Bronn would cast an even wider net, often citing several hundred fossil catalogs and monographs as the basis for his analyses. What makes Bronn's *Italiens Tertiär-Gebilde* especially noteworthy, however, is the extent of the numerical analysis that follows the catalog of taxa, including more than 50 pages of ''General Remarks'' about his faunal analysis, and the inclusion of 26 fold-out tables in which the numerical data were presented. The tables are very much in the botanical arithmetic tradition, being concerned mostly with calculating ratios of certain groups to one another, arranged in succession of geological eras. For example, he initially calculated the relative proportions of genera and species of all fossils animals found in the Italian Tertiary, which in subsequent tables he compared with data from catalogs from other regions (Figure 3). Further tables exhibit values for the absolute number of groups present in the Italian Tertiary, the absolute number of extinct groups, and the percentage of extinct groups present in each formation. These calculations are very similar to those performed for the French Tertiary by Lyell and Deshayes. Finally, Bronn focused on relationships among genera and species of one group – the gastropods – in order to calculate the relative 'richness' of particular taxa during each Tertiary formation (the prevalence of one taxon versus another), and to trace the change in this relationship over time.

Tabelle II. Zu Seite 144.

Klassen und Ordnungen nach Lamarck.	I. Gesammtzahl der				II. Uebergangs- bis Kohlen-Gebirge.				III. Flötz- bis Jura-Gebilde.				IV. Kreide-Formation.				V. Tertiär-Bildungen.			
	Geschlechter.		Arten.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.	
	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.	absolut.	relativ.
Cephalopoden	41	0.172	381	0.188	8	0.033	50	0.025	9	0.038	195	0.096	12	0.050	86	0.042	16	0.067	50	0.025
Trachelipoden	(79)	(0.332)	(656)	(0.324)	(15)	(0.063)	(46)	(0.023)	(11)	(0.059)	(35)	(0.017)	(21)	(0.088)	(34)	(0.017)	(72)	(0.302)	(541)	(0.267)
Zoophagen	38	0.160	350	0.173	4	0.017	13	0.006	5	0.021	10	0.005	9	0.038	10	0.005	35	0.147	317	0.157
Phytiphagen	41	0.172	306	0.151	11	0.046	33	0.017	9	0.038	25	0.012	12	0.050	24	0.012	37	0.155	224	0.110
Gasteropoden	13	0.055	82	0.040									3	0.013	5	0.002	12	0.050	77	0.038
Pteropoden	2	0.008	3	0.001													2	0.008	3	0.001
Conchiferen	(96)	(0.403)	(833)	(0.431)	(38)	(0.159)	(167)	(0.083)	(33)	(0.138)	(156)	(0.078)	(51)	(0.214)	(236)	(0.117)	(58)	(0.244)	(314)	(0.155)
Dimyarier	50	0.210	367	0.181	15	0.063	36	0.018	17	0.071	46	0.023	21	0.100	55	0.027	44	0.185	230	0.114
Monomyarier	26	0.109	230	0.114	10	0.042	16	0.008	14	0.059	62	0.031	18	0.076	80	0.040	11	0.046	72	0.035
Brachiopoden	15	0.063	254	0.125	12	0.050	114	0.056	2	0.008	48	0.024	5	0.021	80	0.040	3	0.013	12	0.006
Sphaeruleen	5	0.021	22	0.011	1	0.004	1	0.001					4	0.017	21	0.010				
Cirrhopoden	4	0.017	14	0.007									1	0.004	1	0.001	4	0.017	14	0.007
Anneliden	3	0.013	15	0.008									1	0.004	1	0.001	3	0.013	14	0.007
Summen	238	1.000	2025	0.999	61	0.255	263	0.131	56	0.235	386	0.191	89	0.373	363	0.180	167	0.701	1013	0.500

Tabelle III. Zu Seite 144.

Klassen und Ordnungen nach Lamarck.	I. Gesammtzahl der				II. Uebergangs bis Kohlen-Gebirge.				III. Flötz- bis Jura-Gebilde.				IV. Kreide-Formation.				V. Tertiär-Bildungen.				VI. Tertiär-Bildungen (ohne *Plastic-clay*) gesondert in							
																					London-clay.				*Obere Meeresformat.; Crag. Süßwasserformat.*			
	Geschlechter		Arten.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.		Geschlechterzahl.		Artenzahl.	
	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.	absol.	relat.
Cephalopoden	14	0.093	317	0.202	5	0.033	49	0.031	3	0.020	150	0.095	7	0.047	104	0.066	3	0.020	9	0.006	3	0.020	9	0.006				
Trachelipoden	(48)	(0.320)	(399)	(0.254)	(12)	(0.080)	(30)	(0.019)	(17)	(0.113)	(76)	(0.048)	(15)	(0.100)	(36)	(0.023)	(44)	(0.293)	(260)	(0.165)	(33)	(0.220)	(158)	(0.100)	(22)	(0.146)	(95)	(0.060)
Zoophagen	23	0.153	184	0.117	2	0.013	5	0.003	4	0.027	9	0.006	3	0.020	8	0.005	24	0.160	167	0.106	19	0.127	110	0.070	8	0.053	50	0.032
Phytiphagen	25	0.167	215	0.137	10	0.067	25	0.015	13	0.086	67	0.043	12	0.080	28	0.018	20	0.133	93	0.059	14	0.093	48	0.030	14	0.093	45	0.028
Gasteropoden	11	0.073	53	0.034	1	0.007	2	0.001	6	0.040	15	0.009	2	0.013	7	0.004	8	0.053	27	0.017	5	0.033	14	0.009	8	0.053	12	0.008
Pteropoden																												
Conchiferen	(72)	0.487	(762)	(0.485)	(20)	(0.133)	(98)	(0.062)	(50)	(0.333)	(326)	(0.207)	(43)	(0.280)	(183)	(0.116)	(42)	(0.293)	(158)	(0.100)	(32)	(0.214)	(71)	(0.045)	(23)	(0.153)	(73)	(0.046)
Dimyarier	43	0.287	360	0.229	9	0.060	13	0.008	31	0.207	155	0.099	23	0.147	70	0.044	36	0.240	122	0.078	25	0.167	56	0.036	20	0.133	57	0.036
Monomyarier	21	0.140	235	0.149	5	0.033	11	0.007	14	0.093	120	0.076	17	0.113	72	0.046	7	0.046	35	0.021	6	0.040	14	0.009	3	0.020	16	0.010
Brachiopoden	8	0.053	167	0.106	5	0.033	73	0.046	5	0.033	51	0.032	3	0.020	41	0.026	1	0.007	1	0.001	1	0.007	1	0.001				
Sphaeruleen	1 Amplexus ?			0.001	1?	0.007	1?	0.001																				
Cirrhopoden	2	0.013	6	0.004									1	0.007	2	0.001	2	0.013	4	0.062					2	0.013	4	0.062
Anneliden	2	0.013	35	0.022	1	0.007	1	0.001	2	0.013	16	0.010	2	0.013	15	0.009	1	0.007	2	0.001	1	0.007	2	0.001				
Summen	150	1.000	1572	1.001	39	0.260	180	0.114	78	0.520	583	0.370	70	0.460	346	0.220	100	0.679	460	0.291	74	0.494	254	0.161	55	0.366	184	0.117

Figure 3. Bronn's early quantitative analysis of Tertiary fossils (Bronn, 1831, Tables II and III)

Rudwick has commented rather dismissively that Bronn's quantitative approach in *Italiens Tertiär-Gebilde* "shows no theoretical structure underlying his 'statistical' analysis," and remarks that "Bronn gives the impression of having had more figures than he knew how to handle, and of not having had any clear theory that he wished to test" (Rudwick, 1978, p. 236). While somewhat harsh, Rudwick's assessment may be accurate for this particular work. *Italiens Tertiär-Gebilde* was Bronn's first – and very early – foray into quantitative analysis of the fossil record, but it represents a step towards a larger quantitative project that I argue *was* genuinely theoretical. In fairness to Bronn, his numerical analysis did yield some tentative conclusions, for example that earlier formations contain a greater percentage of extinct groups than younger ones, which confirmed Lyell and Deshayes' finding, and more importantly gave a precise *quantitative* measure of species richness and replacement (Bronn, 1831, p. 148). It should also be kept in mind that Bronn and his contemporaries did not assume that organisms evolved, so they lacked a theoretical framework in which gradual extinction and replacement made intuitive sense. In fact, it was precisely on the basis of studies like Bronn's and Lyell's that Darwin (who read and corresponded with Bronn) was able to see the fossil record as a record of extinction and evolution. Finally, Bronn himself acknowledged repeatedly that his conclusions were tentative and based on incomplete data. While he asserted that "the relationship of one Tertiary basin with another can be expressed mathematically, if one could assume to know every fossil species in the area well," he was well aware that "these studies are based on very poor foundations" (Bronn, 1831, pp. 155 and 174). Nonetheless, he was prepared to argue that his method offered promise for future studies, concluding that "these studies are sufficient not only to settle a dispute concerning the Italian Tertiary structure, but also to demonstrate the application of a numerical approach to characteristics of the fossil deposits in rock strata, that has so far not been considered" (Bronn, 1831, p. 174).

I argue that Bronn's vision was to treat the fossil record as a record of data, and to subject those data to numerical analysis revealing patterns of life history that could be reduced to principles or even laws. In later works, such as his comprehensive survey of the fossil record *Lethaea Geognostica*, and his mammoth, multi-volume catalog *Index Palaeontologicus*, Bronn compiled a global bibliographic survey of the entire known fossil record (Bronn, 1835, 1848). These efforts were differentiated from other contemporary catalogs in that they attempted to assemble a collection of data as a basis for testing hypotheses about the fossil record using numerical analysis, and for deriving laws of nature from the patterns that resulted. This was, to put it mildly, a challenging enterprise for a mid-nineteenth century paleontologist: Bronn lacked the technology for easily storing and retrieving his data that paleontologists would possess a hundred years later with the advent of computers, and he was also working in an era before the development of sophisticated statistical techniques involving mathematical probability theory that permitted tests of significance, analysis of bias, and correlation of causal relationships in the data being analyzed (Porter, 1986, pp. 3–5). It would be anachronistic to say that Bronn was engaged in the same project as the later twentieth century paleontologists'; however, what Bronn's work does demonstrate is the emergence of a new epistemic value that saw the fossil record as a collection of data to be subjected to quantitative analysis.

Bronn, who for many years held a professorship in natural science at Heidelberg, was considered the leading German paleontologist of his day. Interestingly, his academic training was partly in "cameral studies," which may be where he acquired his interest in elementary statistics and numerical tabulation (Gliboff, 2007, p. 63).[8] Today, Bronn is chiefly

[8] Cameralism, or the science of public administration, was an important influence on northern European naturalist thought during the eighteenth and nineteenth centuries, as Lisbet Koerner has demonstrated in the case of Linnaeus (Koerner, 1999).

remembered for his herculean efforts at compiling knowledge of the fossil record. His student Karl A. von Zittel (who himself became a leading German paleontologist) later characterized Bronn's *Lethaea Geognostica* as ''the first attempt at a Chronological Succession of fossil organisms,'' and called it ''a masterpiece of scholarship'' that ''summarizes all that was previously known about stratigraphy and palaeontology.'' Likewise, Zittel cited the ''great influence on the development of palaeontology'' that the volumes of the *Index Palaeontologicus* exerted, which ''were for several decades the chief books of reference for all the more comprehensive palaeontological works'' (Zittel, 1901, pp. 364–365). But Bronn was also greatly interested in using paleontology to interpret the history of life, and he laid out a grand theory of laws of geological and organic development in several works, including the three-volume *Handbuch einer Geschichte der Natur* and the more concise *Untersuchungen über die Entwicklungs-Gesetze der Organischen Welt* (Bronn, 1841, 1858). Bronn's theory of development was not evolutionary, but he did correspond with Darwin and was appreciative enough of Darwin's work that he personally translated the *Origin of Species* into German immediately after it was first published (Gliboff, 2008). While Bronn's ideas never achieved the success of Darwin's, an essay that was the genesis for the *Entwicklungs-Gesetze* was awarded the Grand Prize of the Paris Academy of Sciences in 1857, and he is rightly considered one of the first important theorists of the modern professional discipline of paleontology.

A major feature of both the *Geschichte der Nature* and the *Entwicklungs-Gesetze* are the inclusion of hundreds of pages of tables of numerical data. These tables list, count, and arrange fossil species by taxonomy, stratigraphy, and geography, as well as provide calculations of their ratios and relative distributions over geological time (Gliboff, 2007, p. 268). I will limit my discussion here to the *Entwicklungs-Gesetze*, since that work is more focused on organic development (*Geschichte der Nature* considered the formation of the solar system and the earth, as well as the history of life), and being later it is the most mature presentation of Bronn's views; however, both works take essentially the same approach to numerical analysis. The *Entwicklungs-Gesetze* does not include a fossil catalog, but Bronn explained that his data were drawn from his *Index Paleontologicus* and other published sources. Roughly 60 pages of numerical tables occupy the first part of the book, followed by some 400 pages of analysis and discussion. While many of the tables resemble those of the *Italiens Tertiär-Gebilde*, in the *Entwicklungs-Gesetze* the numerical tables are merely the starting point for the analysis, rather than an end in themselves. This is not mere "Tabellenstatistik"; it is a genuine attempt at using numerical analysis to derive regularities and laws from the fossil record. As Sander Gliboff puts it, Bronn "makes a point of grounding his laws of historical change in his fossil data, and denying their derivability from the study of living organisms alone" (Gliboff, 2007, p. 263). Gliboff explains that Bronn was motivated by a sense of proper *Wissenschaftsliche* method that privileged systematic organization of empirical data, if possible using quantitative techniques, and that attempted to reveal general laws of nature without any preexisting theoretical assumptions. This approach did not necessarily attempt to derive the *causes* of these laws (although Bronn explicitly rejected theological explanations), but rather "most of the time, he claims no more for his 'laws' than that they are abstracted from repeating patterns in the data" (Gliboff, 2007, p. 270). It is therefore an extension of Humboldt's ideal for science, but in practice goes well beyond simple "botanical arithmetic." Bronn's methodology also epitomizes one of the putative central features of data-driven science: that hypotheses are to be derived from data via induction, and not superimposed at the outset.

Bronn began the *Entwicklungs-Gesetze* by discussing the limitations in our knowledge of fossils, and the challenges of ever fully reconstructing a history of past life (Bronn, 1858, p. 3). These sorts of apologies for the fossil record were becoming more common at this time, and indeed have been a staple of paleontological literature ever since. Nonetheless, he expressed confidence that the fossil record was a source of legitimate and valuable information about the history of life, and that efforts such as his own *Index Paleontologicus* had helped improve knowledge by rendering the empirical data in a more systematic manner (Bronn, 1858, p. 7). In the book's second section, "Working out the Problem," Bronn used a striking material metaphor to describe the fossil record that, I argue, illuminates some of contemporary paleontologists' central epistemic concerns regarding the status and nature of the fossil record. It is worth quoting at length:

> The earth crust is a great book, her layers are the leaves of the same, fossils, the letters of the alphabet with which it is written, and the contents are the story of creation, of which no living eyewitness can give news. But those pages are incomplete, broken, jumbled up and faded before us; we need to organize them and to search to supplement what is missing; some gaps can be restored by drawing from other places; the interpretation finds wide scope and the

> discovery of new fragments, which have been missing, often makes the emendation of earlier patchings [*Einflickungen*] necessary. The alphabet, in which the book was written, was long unfamiliar to us; we had misunderstood it and began first to decipher and comprehend it as we began to look for the key in our present nature; we recognized with astonishment the language of our present, and to see that that the laws in both are the same, and only the characters in the alphabet gradually changed somewhat. The author of this book has the greatest credibility, because he was contemporaneous with the events he describes to us; he was the architect of our earth crust itself, which many events of the time through autobiotype [*Autobiotypie*] have been artistically represented. We receive from it a more or less complete knowledge of the creatures, which existed at the time, of their number and organization, of the laws by which they have been distributed in time and space, of the order in which they followed each other.... Not easily has any significant event in the history of the earth's surface occurred, that to us would not be betrayed by the type, the condition, the association, and the change of the fossil remains. (Bronn, 1858, pp. 75–76)

This metaphor of the fossil record as a "book" or "text" was a fairly common trope at the time, used by both Darwin and Lyell, among others. It is possible that Bronn borrowed his imagery from Lyell, who had employed the metaphor in his *Principles of Geology*, but the specific language Bronn used is much closer to Darwin's subsequent discussion in the *Origin*, where Darwin explained that

> I look at the natural geological record, as a history of the world imperfectly kept, and written in a changing dialect; of this history we possess the last volume alone, relating only to two or three countries. Of this volume, only here and there a short chapter has been preserved; and of each page, only here and there a few lines. Each word of the slowly-changing language, in which the history is supposed to be written, being more or less different in the interrupted succession of chapters, may represent the apparently abruptly changed forms of life, entombed in our consecutive, but widely separated formations. (Darwin, 1859, pp. 310–311)

The significance of the metaphor is in what it demonstrates about how Bronn – and perhaps other paleontologists – understood the nature of the fossil record. Philosopher Linda Patrik has argued that there are two very different meanings of "record" that have informed the historical sciences. In the first case, records are "static, physical things that are the causal effects of what they record,'' while in the second, a record is ''a text, comprised of material symbols that signify what they record'' (Patrik, 2000, p. 123). Patrik explicitly characterizes the fossil record as an example of the former sense, which she likens to a phonograph record, and which she describes as '''passive' in the sense that it records its causes by preserving the static effects of these causes; its recording occurs simply because it bears an imprint.'' To this she contrasts the more complex semiotics of 'textual' records, such as human documents or artifacts, which she argues are comprised of ''a body of signs that encode ideas and information about past events,'' and which ''may lie, exaggerate, or mask the truth, either through deliberate or unconscious choice on the part of the author'' (Patrik, 2000, p. 124). According to Patrik, modern paleontology has adopted a model where fossil evidence is ''a physical record, not a textual record'' (Patrik, 2000, p. 124). However, I think that Lyell's, Darwin's, and Bronn's explicit use of a metaphor where the fossil record is described as a text – a practice that continued well into the second half of the twentieth century – challenges Patrik's interpretation. I argue that the fossil record can be both kinds of records simultaneously: one that passively records physical effects (impressions of once-living organisms), and also a text containing signs that encode information about past events. As Bronn's quotation makes clear, the epistemic position of the interpreter to the text is not straightforward: The text is fragmentary and incomplete and thus 'misleading,' since it was written in an alphabet that needs to be decoded. They key transformation in Bronn's formulation, though, is that the ''alphabet'' of this book is decodable through *numerical* analysis; it thus encodes knowledge that can only be recovered using a hermeneutics of data analysis that reveals underlying patterns and laws inscribed by an ''author'' (though again, Bronn's author is nature itself, not a deity).

The development of a quantitative-hermeneutic approach to the fossil record focused on data is central to the epistemological convention Bronn helped shape, and has continued to have influence throughout the twentieth century. From the earliest fossil compendia, the fossil record has always been considered a text, but different hermeneutic approaches have characterized different readings of that text. Late eighteenth and early nineteenth century approaches were particularly concerned with a visual imagery in which fossils were depicted either in situ in the earth's layers (which supported the notion that the 'leaves' of the text being interpreted were geological strata), or disembodied in the specimen drawer, where the visual imagery primarily offered clues about the morphology and function of organic structures and systematic relationships between different groups of fossil organisms. By the mid-nineteenth century, however, the text became the history of life, and the new hermeneutics that emerged saw fossils as data points for quantitatively interpreting patterns and laws that governed its development and unfolding. Although I will not focus on evolutionary theory here, one obvious example of this is the way that fossil data became incorporated as a resource for understanding phylogenetic patterns, and the fossil record 'became' a record of evolutionary change (Sepkoski, 2012). Thus, the ‘‘fossil record’’ could simultaneously be many different kinds of texts (the earth itself, a pictorial catalog, a list of taxa, tables of numerical data), but each text – whether ‘‘natural’’ or man-made (e.g., a physical book) – required different hermeneutic strategies for its interpretation.

Bronn's approach to decoding the text primarily involved numerical analysis of data, but it also included images as well. These images were not illustrations of fossils, but rather diagrams that contributed to a new visual genre in paleontology. One prominent example of this visual genre is images that represented the longevity and diversity of groups of organisms over time, which in modern paleontological parlance are called ‘‘bubble’’ or ‘‘spindle’’ diagrams, and are essentially visual summaries of quantitative analysis. Bronn is among the first paleontologists to present such diagrams, although some contemporary examples do exist.[9] In the later analytical section of the *Entwicklungs-Gesetze*, Bronn used spindle diagrams to demonstrate the gradual appearance and disappearance of groups from the fossil record as evidence against sudden, catastrophic faunal change (Figure 4). As he explained, in reference to one figure, ‘‘the thickness of the horizontal lines reflects the strength of the development of the [taxa discussed] above, and of other major orders and families, by reflecting the number of genera in a family’’ (Bronn, 1858, p. 312). The advantage of such diagrams is that they provide an immediate visual summary of the relative patterns of development among the various groups, distilled from the lengthy tables of data that record the first and last appearances of taxa in the fossil record. Bronn generated these diagrams for a variety of groups of organisms at different levels of taxonomic resolution, and they served as the basis for a variety of conclusions about the history of life: that species have appeared and become extinct continuously; that species'

[9] Both Louis Agassiz and Richard Owen occasionally used spindle diagrams to depict, respectively, the historical diversity of fish and reptiles (Archibald, 2009, pp. 586–588).

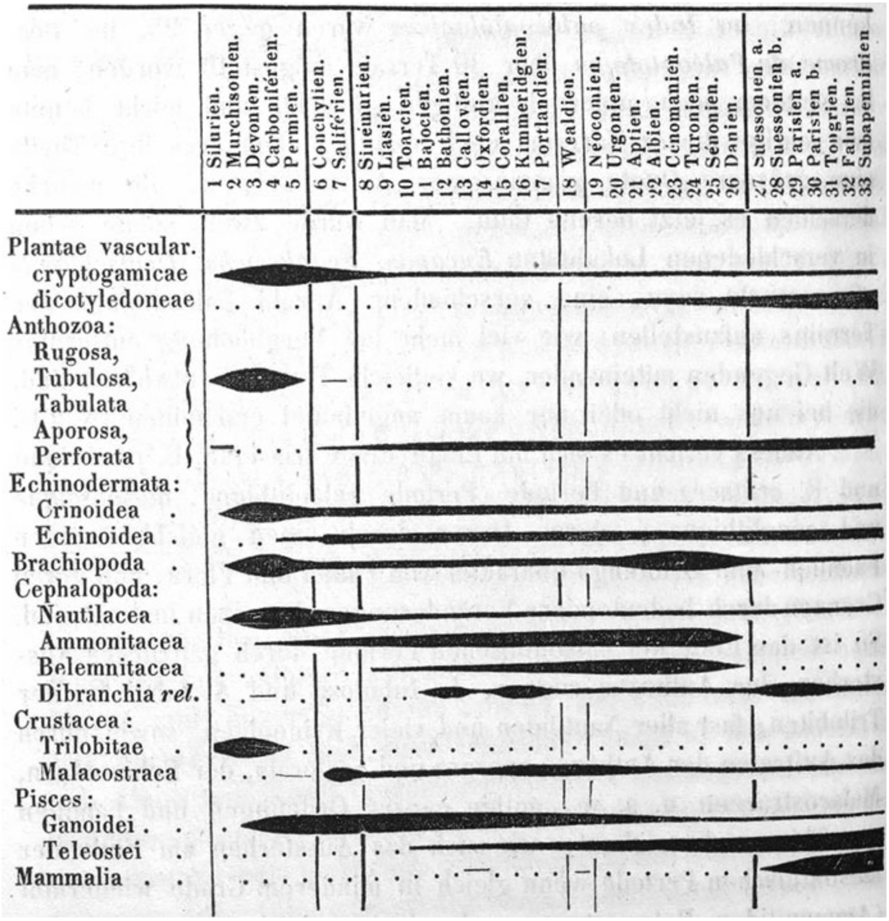

Figure 4. A "spindle" diagram in Bronn's *Untersuchungen über die Entwicklungsgesetze der Organischen Welt*. The length of each line represents the longevity of the group, while the thickness the relative within-taxon diversity (Bronn, 1858, p. 312)

durations do not exactly match geological periods and often cross stratigraphic boundaries; and that species persist on the average for a very long time. Overall, the pattern of the history of life that Bronn interpreted from his data was one in which, over the *longue durée*, certain groups appear, flourish (become thicker in the spindle diagrams), and then gradually wane as other groups rise to take their place. This resulted, he believed, not from evolution or natural selection, but from a progressive law of development, about whose cause Bronn refused to speculate – it is an ‘‘empirical law’’ drawn from the patterns of data.

Bronn was not the only paleontologist of his era to approach the fossil record in this way, but his exploration of the fossil record as a record of data was among the most ambitious. As we have seen, Lyell made early attempts at drawing generalizations about life history from statistical analysis of data, and even Darwin based some of his conclusions about geological succession in the *Origin* on numerical studies of the rates of species replacement by Bronn and F.J. Pictet (Darwin, 1859, pp. 312–315). One of the most ambitious contemporary attempts at extrapolating patterns of organic history from fossil data was John Phillips' *Life on the Earth* (1860), which was published two years after Bronn's *Entwicklungs-Gesetze*. As a young man, Phillips had contributed to the project of establishing a general stratigraphy by collecting and cataloging fossils in different regions of England, but *Life on the Earth* was an attempt to draw larger conclusions from a more global database of fossils.

Unlike Bronn, Phillips saw organic and geological history as part of a divine, providential plan, but like Lyell, Bronn, and Darwin, he described the fossil record as a text: ‘‘The Book of the Strata, inscribed with the earlier Wonders of Nature, has been given to [mankind to] be opened with care and deciphered with reverence, by the help of comparison with the living inhabitants of the Land and Sea’’ (Phillips, 1860, p. viii). As Phillips saw it, ‘‘Standing by the stream of life, we have surveyed the variations in its course, and appealed to history and experience, for the data which might guide us to a right view of its incessant fluctuations, and its recurring uniformities’’ (Phillips, 1860, p. 1). Drawing primarily on the second edition of Morris's *Catalogue of English Fossils*, Phillips set about counting the various marine species discovered in British strata over the whole of the lower Paleozoic, from which he calculated the relative numbers of species and higher taxa in each period. While this count was fairly straightforward, Phillips did recognize that the ‘‘Book of the Strata’’ could not always be read at face value: he was well aware that not all organisms would have been preserved with equal frequency, and that not all stratigraphic units were of equal thickness. For this reason, he corrected his account by estimating the number of species per unit thickness of sediment, which he believed would give a more reliable overall estimate of the history of the diversification of life (Phillips, 1860, p. 60). This is one of the earliest examples of ‘‘bias correction’’ in the history of analytical paleontology,

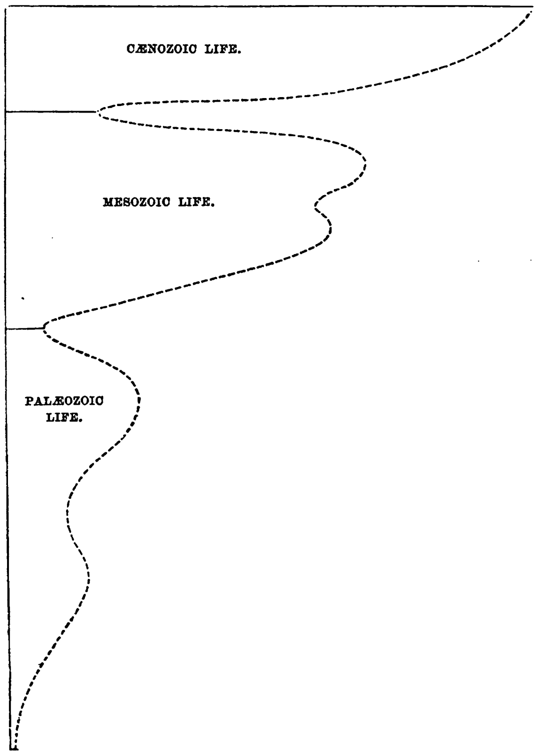

Figure 5. Phillips' diversity curves for the three stages of prehistoric life. Note that the figure is oriented vertically, with time beginning at the bottom of the graph (Phillips, 1860, p. 66)

and it is an approach which would (though with increasingly greater sophistication) become central in later twentieth-century paleontology.

From this numerical analysis, Phillips made a broad generalization about the history of life: the rates of change of the composition of fauna have not been constant over geological time, and the history of life may be grouped into three distinct eras, where "each of the characteristic and prevalent fauna begins at a minimum, rises to a maximum, and dies away to a final minimum, to be followed by another system having similar phases" (Phillips, 1860, p. 64). Phillips then represented this pattern with a striking graph depicting three successive diversity curves, corresponding to what he considered the three great periods of life, the Paleozoic, the Mesozoic, and the Cenozoic (Figure 5). He also performed a number of interesting additional analyses, but I will draw attention to two further distinct visual summaries of his data. In the first

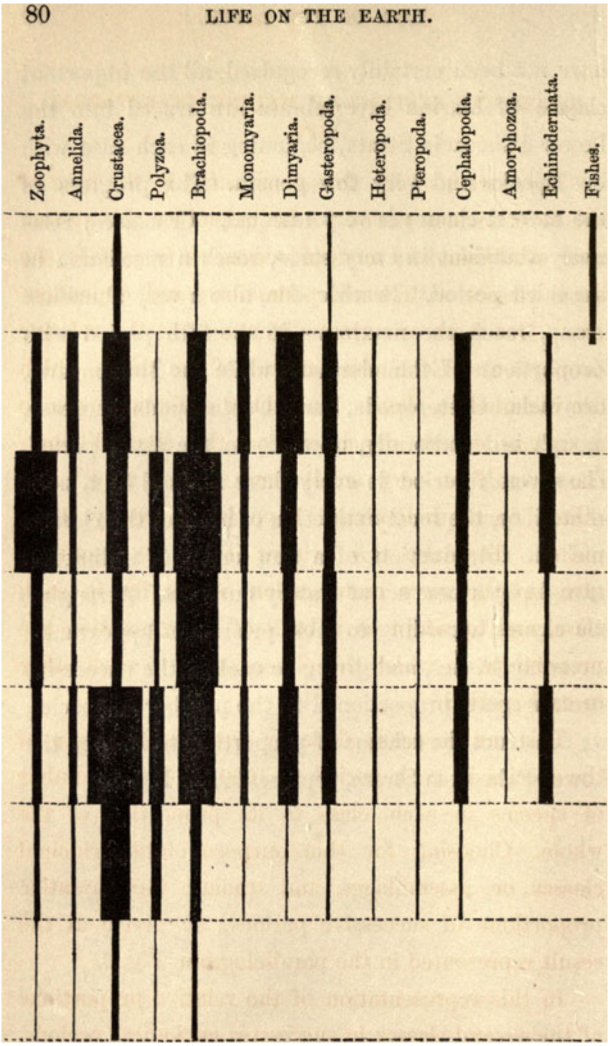

Figure 6. Phillips' spindle diagrams for major marine classes (Phillips, 1860, p. 80)

example, in order to visually convey the representative dominance and succession of individual classes of organisms in each period, Phillips presented a "spindle diagram" depicting "a scheme of proportionate life for the lower Palaeozoic strata," very similar to the ones Bronn published in the *Entwicklungs-Gesetze* (Figure 6). Here, the relative thicknesses of the individual lines give a visual demonstration of their relative dominance during different geologic eras (geologic time begins at the

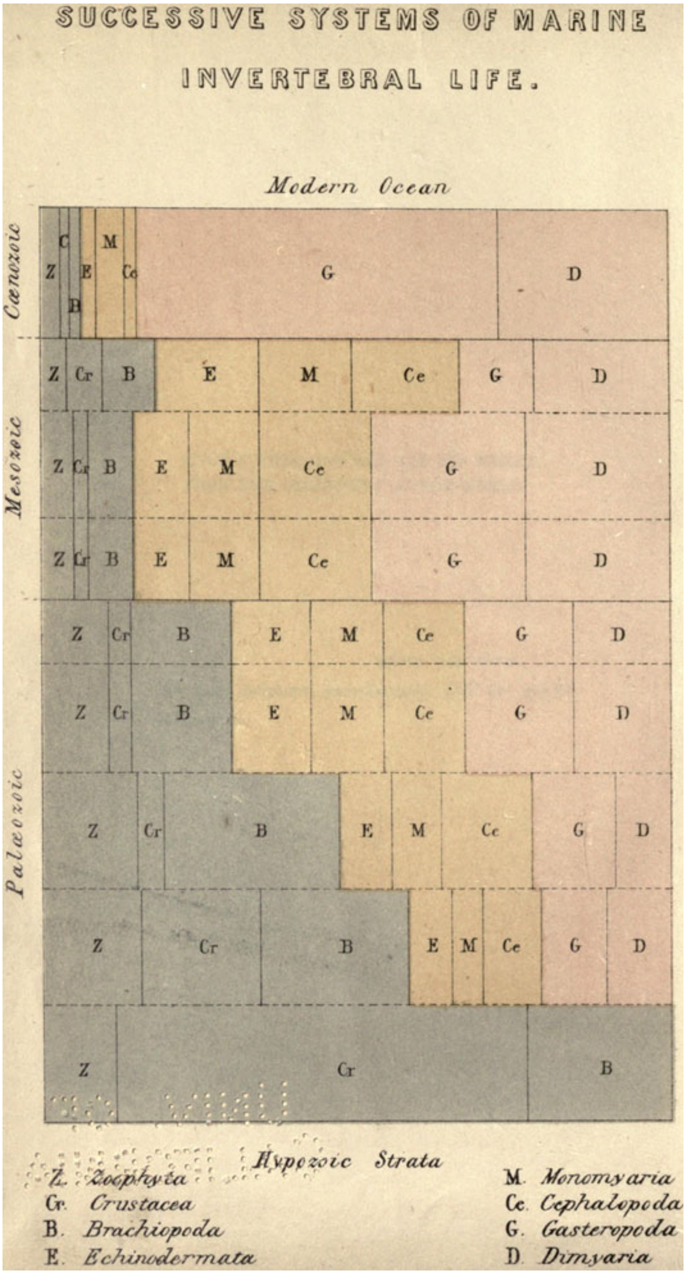

Figure 7. The frontispiece to Phillips' *Life on the Earth*, representing "the relative proportions of the several classes in successive geological periods." See text for explanation (Phillips, 1860)

bottom of the figure). And, in a final image, which was reproduced in color as the frontispiece of the book, he prepared a figure showing ‘‘the relative proportions of the several classes in successive geological periods,’’ in which different tints were used to indicate the broader directional patterns of diversity of each of eight major classes across geological time. As Phillips explained, ‘‘by the blue tint [left] those classes which suffer diminution with time, and by the red tint [right] those which from small beginnings grow to great preponderance, while the yellow tint [center] is assigned to classes which scarcely appear in the early period, but swell out in the middle of the scale so as to equal or overmatch either of the other classes’’ (Figure 7). These three images are, individually, representations of the fossil record in pictorial schematic form; the fossils have been reduced to numerical data, and the data have been translated into visual images. ‘‘Thus appears in a striking light,’’ as Phillips concludes, ‘‘the great difference between the systems of oceanic life in earlier and later periods, the nature of this difference, and something of the method of variation which binds the whole into one plan, and connects the dawn of created life with this our breathing world’’ (Phillips, 1860, pp. 81–82).

Some of the themes that emerge in Bronn’s and Phillips’ approach to the fossil record, then, are a reconceptualization of the history of life as a history of data, the desire to make paleontology more ‘‘scientific’’ by making it quantitative, and the expression of patterns and regularities in the data using new visual forms. These motivations were tempered by some inherent limitations, however, in the quality and availability of the data itself, and in technologies and techniques for handling and analyzing the data. While the basic epistemic concern was fairly well-articulated, the means of fully expressing it would not be available for at least another hundred years.

Towards a Natural History of Data

From the beginning of the nineteenth century to the 1860s, we have seen the emergence of an approach to the fossil record in which what was initially a pictorial catalog of objects was transformed to a repository of data. That is not to say, as I have stressed, that the data-oriented approach to the fossil record *replaced* the pictorial one. As Daston and Galison have pointed out, ‘‘Epistemic virtues do not replace one another like a succession of kings. Rather, they accumulate into a repertoire of possible forms of knowing’’ (Daston and Galison, 2007,

p. 113).[10] However, during the twentieth century, epistemic virtues of quantification, statistical analysis, and graphical abstraction of data patterns became more and more central in that branch of paleontology known as paleobiology (Sepkoski, 2012). By the late twentieth century, this ''paleobiological'' approach had become ascendant in paleontology, and this shift brought with it a final transformation of the fossil record to a digital database. This final transformation of the record – in which the compendium of fossil data was transferred into the medium of the computer (initially through punch-cards and eventually magnetic storage devices) – was less a shift in a conception of what the fossil record was, than an opening of possibilities for what one could do with it. In many ways, the epistemic concerns of late-twentieth century paleontologists have not been so different from Bronn's or Phillips'; rather, translation into the digital domain has allowed paleontologists to express and explore those concerns in ways that, prior to the advent of computer technology, was simply not possible. Another way of putting it is that while computers have certainly enabled new data practices, the twentieth-century computer revolution did not create a scientific culture oriented towards data.

In fact, basic paleontological data practices of have not changed so radically in the more than 150 years since Bronn published his works. When, in the mid-1960s, a committee of paleontologists met under the auspices of the Geological Society and the Palaeontological Association in Britain attempt to reform data practices in paleontology, and to produce a new database of the fossil record, the product was a physical book, titled simply *The Fossil Record* (Harland et al., 1967). By the second half of the twentieth century, the quantity of fossil data had multiplied exponentially from what was available in Bronn's day, and standard reference compendia, such as the multi-volume *Treatise on Invertebrate Paleontology*, had already grown to many volumes and tens of thousands of pages (the *Treatise* has been continually updated since the 1950s, and now contains more than 50 volumes). The aim of *The Fossil Record* was ''to assemble, in one volume, through the efforts of many specialists, a useful collection of data'' that would be more tractable for analysis than resources like the *Treatise* (Harland et al., 1967, p. 1). *The Fossil Record* abstracted only one piece of information about fossils: the range data (e.g. the first and last appearances in the fossil record) of the higher taxonomic groups of fossil organisms in the known record, based on existing published literature. The volume did not

[10] This point about the accumulation of epistemic virtues is also made by Pickstone, 2001.

attempt to document the *entire* fossil record, since data about many lower taxa (e.g. family, genus, and species) were not included, so the lists were not nearly as long as those in works like Bronn's *Index Paleontologicus*. Rather, the editors explained that they considered their volume to be "an abstract" of the fossil record itself, believing that "such a compilation would be a valuable initial source of information for a variety of uses, such as the critical evaluation of theories concerning the history of life, and, not least, to draw attention to the deficiencies of our present knowledge" (Harland et al., 1967, pp. 2–3).

While it was not an electronic database, *The Fossil Record* did point towards the utilization of computers for storage and numerical analysis of fossil data – it was, in a sense, a template for an eventual digital database, and an exploration of how one might be used. The volume concluded with an essay on "Numerical Analysis of *The Fossil Record*," in which its authors, J.L. Cutbill and B.M. Funnell, described some initial attempts at translating the published data onto computer tapes for running numerical analyses on a mainframe computer. Their basic approach was quite similar to Bronn's: Cutbill and Funnell tabulated data for 20 major groupings of organisms, observing the relative incidence of each taxon during each stratigraphic stage, and used the computer to generate graphs representing the patterns of change in those data for each group and to superimpose the data for all groups into a single, graphical representation of change in the history of life over time (Figure 8). The authors admitted that their approach – and their data – had many shortcomings, but they contended that "in so far as the patterns they reveal may promote closer investigation of their generating causes a useful purpose will have been served" (Cutbill and Funnell, 1967, p. 793). For these kinds of analyses to be more authoritative, they pointed out, a much more complete and comprehensive fossil database would have to be established.

A few years later, in 1972, a group of American paleontologists met for a weekend of informal discussions at the Marine Biological Laboratory at Woods Hole, Massachusetts, about how to explore novel theoretical approaches to the study of fossil data. This so-called "MBL group" consisted of the paleontologists Stephen Jay Gould, Thomas J.M. Schopf, David Raup, and the theoretical ecologist Daniel Simberloff, and the meeting was organized by Schopf as "a self-conscious attempt to introduce more theory into our mass of facts" (Thomas J.M. Schopf to David Raup, 5 March, 1972. Thomas J.M. Schopf Papers, Smithsonian Institution Archives.). This meeting was part the beginnings of a much larger movement to establish an analytic or

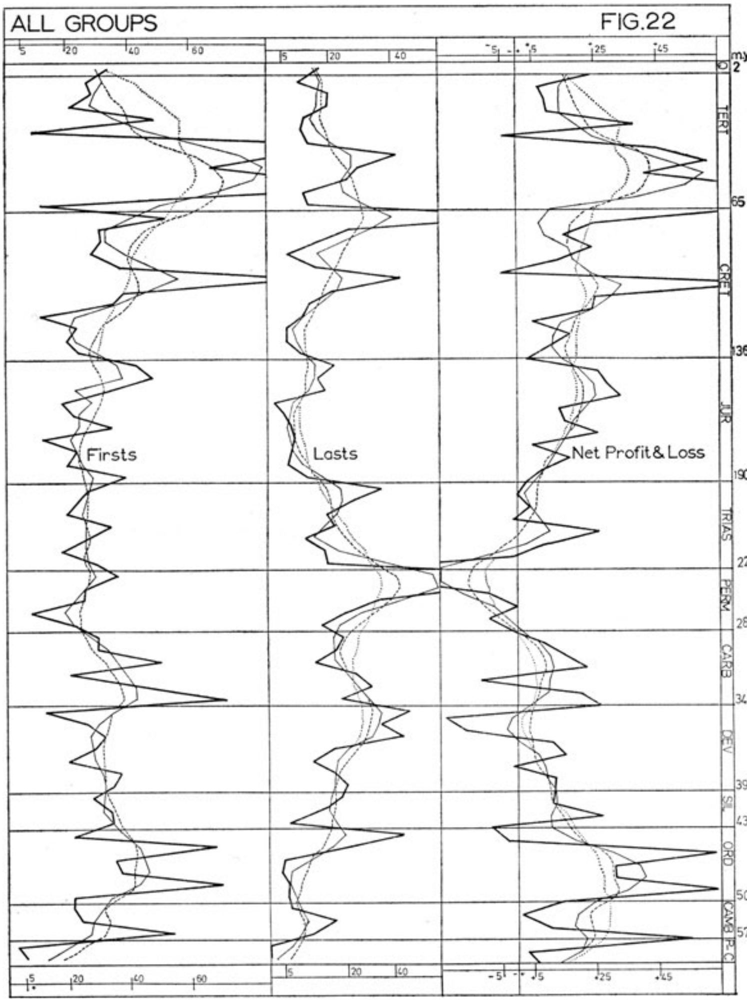

Figure 8. Cutbill and Funnell's quantitative analysis of data from *The Fossil Record* (Cutbill and Funnell, 1967, p. 819)

"nomothetic" paleontology that unfolded over the 1970s and early 1980s, which had the goal of promoting greater deployment of models, numerical analysis, and theoretical innovation in paleontology (Sepkoski, 2012). The salient point here, however, is that the MBL group hoped to perform exactly the kinds of numerical analyses on fossil data as Cutbill and Funnell had explored: Raup remembers that Schopf brought the entire multi-volume *Treatise*, as well as *The Fossil Record*, "and we put it on the table," while Simberloff came equipped with one of the earliest programmable calculators. The question, Raup recalls, was "what can we do that's different?" (Raup, quoted in Sepkoski and Raup, 2009, p. 463). The answer, as it turned out, was not much: the data were determined to be too fragmentary and unreliable to serve as the basis for more sophisticated evolutionary-ecological analysis.

Faced with failure, on the last day of the meeting Raup suggested a radical option: rather than analyzing the actual fossil record, he proposed writing a computer program that simulated the evolution and extinction of hypothetical fossil lineages based on a few simple parameters. Ultimately, this simulation approach (called the "MBL model") met with only limited success, but its history is outside the scope of this paper.[11] However, the ultimate limitations in the MBL modeling approach provided new inspiration for paleontological database projects. One of the major contemporary criticisms of studies of simulated fossil data was that their conclusions were of little comparative value so long as estimates of the actual fossil record remained unreliable. This was evident to Gould as early as 1973, when he set his graduate student, Jack Sepkoski, the task of compiling data on orders, families, and genera from existing compendia of fossils such as the *Treatise* and *The Fossil Record* to serve as the basis for comparison with the MBL simulation runs. Sepkoski began assimilating data from other, additional sources (such as monographic literature) in a project that ended up ballooning to occupy the first decade of his professional career. The product of ten years' examining and compiling all of the available data on marine fossil families was published in 1982, and what is now known as the "Sepkoski Database" became (in its day) the largest and most important data collection for the fossil record in the world (Sepkoski, 1994). It was the ancestor of all of the electronic database projects and collaborations currently used by paleontologists, and the source or inspiration for countless studies of patterns in the fossil record over the past 30 years.

While it may have been greater in scope, however, the Sepkoski compendium project (which eventually came to include marine genera as well) was not terribly different from compilations by Bronn and other nineteenth century paleontologists, at least in a formal sense. Like those earlier works, it was effectively a list of fossil taxa, including only information about the first and last appearances of taxonomic groups in the fossil record, along with bibliographic references to the relevant specimens. Materially, it was hardly different at all: the 1982, 1992, and

[11] For a history of the importance of the MBL model for the development of paleobiology, see Sepkoski, 2012, Chap. 9.

2002 editions of the compendium were published as physical books, although the last edition also included a cd-rom containing the database in electronic format (Sepkoski, 1982, 1992; Sepkoski et al., 2002). Despite the availability of computers and word processors, Sepkoski compiled the compendium by hand, in a series of notebooks he kept over many years. In an autobiographical account, he described updating the compendium by painstakingly slicing entries out of the proofs with a razor blade, and affixing them with scotch tape in their proper places (Sepkoski, 1994). And the kinds of analyses Sepkoski and other paleontologists performed on the data were similar, at least in objective, to those of their nineteenth century predecessors: Sepkoski generated comprehensive spindle diagrams from his data (Figure 9), and his own most famous contribution to understanding the history of life was a series of studies in which he used his compendium as the basis for analysis of the history of faunal diversity of the entire marine fossil record, which he concluded could be represented as a series of three, overlapping diversity curves of roughly logistic shape (Sepkoski, 1984) (Figure 10). This graph is now iconic in modern paleobiology, but it is – as Sepkoski was well aware – extremely similar in general outline to the three-curve graph Phillips published in 1860.

Of course, there are major differences between nineteenth and late twentieth century approaches to the fossil record as well. While the fossil record may still be treated like a "text," the techniques used to interpret that text have changed considerably. One of the most important analytical problems in modern paleontology is the understanding and, as far as is possible, correction of "bias" in the fossil record. This was the same problem that Phillips attempted to resolve: because of inconsistencies in fossil preservation, geographic distribution, and paleontological collection practices, the fossil record is an unreliable document, and cannot be "read" at face value. An array of statistical techniques now exist for detecting and resolving bias, including multivariate statistics, rarefaction, and analysis of variance; these techniques testify to the importance of the numerical (or quantitative) hermeneutics paleontologists use to approach their text. The advent of modeling and simulation have also contributed new meanings to what the fossil record "is," in both an epistemological sense, as the source of experimentation and hypothesis testing, and in an ontological one, as a collection not of things but of "pure data." And the forms of visual presentation of patterns of data derived from the fossil record have proliferated, contributing to a rich visual epistemology of data in which visual hermeneutics plays a key role in interpreting evolutionary patterns.

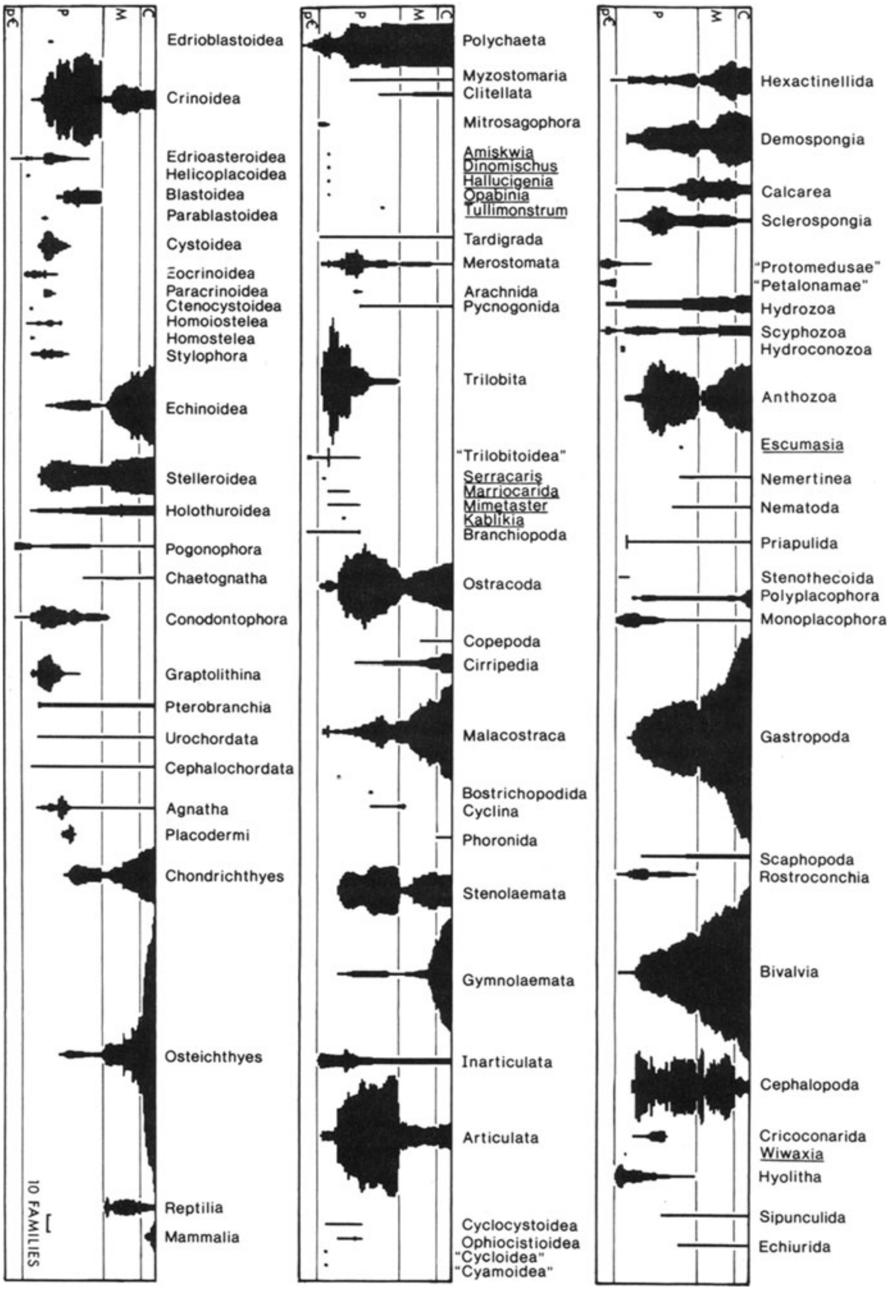

Figure 9. Spindle diagrams from Jack Sepkoski's factor analysis of Phanerozoic marine taxonomic diversification (Sepkoski Jr., 1981, p. 38)

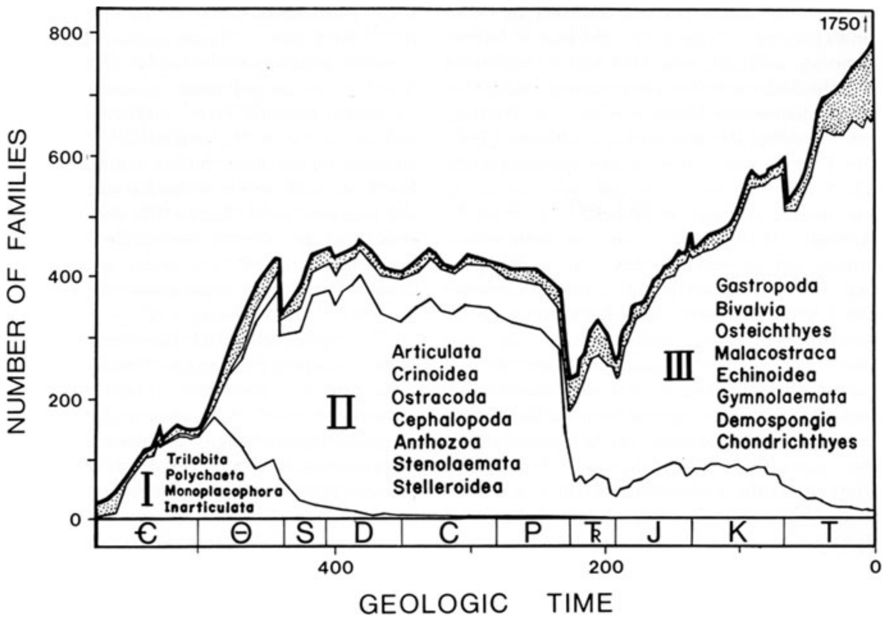

Figure 10. Sepkoski's 3-curve graph of Phanerozoic marine faunal diversity, from Sepkoski, 1984, p. 260

The relationship between the visual and numerical hermeneutics used by paleontologists has sometimes been complex: one illustration is an episode during the early 1980s, when Sepkoski and Raup were performing some of their initial analyses of the Sepkoski compendium. Raup used a word processor to transfer portions of the database onto a computer, which he sorted to examine which groups went extinct during which intervals of time. The result of this experiment was a "remarkable graph" showing several major extinction peaks that were clearly distinguishable from "background" rates of extinction. As Raup later explained, "We were looking at the computer output mostly as a series of pictures – looking for a gestalt that could lead us in interesting directions" (Raup, 1986, p. 114). This graph – which had been generated automatically by the computer – became the source of a new understanding of the role of mass extinction in the history of life, including the hypothesis that mass extinctions followed a 26 million-year "periodicity" triggered by asteroid or comet impacts (Raup and Sepkoski, 1984). However, Raup did not immediately trust this visual gestalt, and it was only after subjecting the pattern to rigorous statistical tests that Raup and Sepkoski felt confident enough to assert the validity of the pattern. Debates about periodicity in extinction – and the veracity of these kinds of visual projections of patterns in the fossil record more generally – have hinged on arguments about the reliability of the statistical tests used to verify them. Paleontologists now often ask whether patterns derived from quantitative analysis of the fossil record are a genuine biological "signal," as opposed to an artifact of sampling or other kinds of bias, and whether that "signal" matches other kinds of data, such as the chronology established by molecular phylogenetics (the "molecular clock") (Benton and Emerson, 2007). The language and substance of these debates reinforces the sense in which the fossil record has been converted to a record of pure data – a 'digital signal' – and the study of the fossil record has become, as Jack Sepkoski used to sometimes half-jokingly describe, "a natural history of data" (Foote, 1999, p. 235).

Conclusion

As Geoffrey Bowker has put it, "Perhaps the most powerful technology… in our control of the world and each other over the past two hundred years has been the development of the database" (Bowker, 2005, p. 108). As I have argued in this paper, the emergence of databases in paleontology is continuous with a history of information management and analysis stretching back to the mid-nineteenth century. It is, in a sense, a history of the relationship between objects and knowledge, as seen through an examination of the practices and technologies that scientists use to re-present the world as data. Contrary to what we might have expected, this process began long before the invention of the technologies that characterize the modern "data-driven sciences," such as the computer and the internet. The epistemic concerns that motivate paleontological analysis of the fossil record today have a clear genealogical relationship to those that motivated nineteenth century naturalists like Lyell, Bronn, and Phillips, though the technologies that support the compilation, storage, and analysis of paleontological data have changed significantly.

The foregoing analysis suggests several observations about the development of data practices in paleontology, and about the relationship between natural history collections and databases. First, paleontologists' epistemic concerns have been embodied in the material culture and practices by which they represent their data. Daston and Galison have explored how compilations of scientific images are sites for the inscription of "epistemic virtues" that change over time (Daston and

Galison, 2007, p. 42). However, as we have seen, not all representations of the fossil record involved pictorial images. Daston and Galison's point about the relationship between images and epistemic virtues can be expanded more broadly to include practices and forms of representation (the list, the table of data, the graph, and even the metaphor) that are not necessarily pictorial. The material forms which these representations took have shaped the epistemic concerns of paleontologists. As Bruno J. Strasser has noted, databases are not merely repositories of information, but rather "tools for producing knowledge" (Strasser, 2011, p. 63). Printed compendia and tables of data have had a similar function in shaping the kinds of knowledge paleontologists produced. This leads to a second point, about the relationship between technology and data. Paleontologists' methods of representing and analyzing the fossil record have always depended crucially on available technologies. These technologies included tools for visual depiction, statistical and mathematical techniques, and instruments for storage and analysis of information. However, the relationship between technology and epistemic concerns has not been unidirectional. In some cases, epistemic values regarding the management and analysis of data preceded the availability of adequate technologies to fully deploy them. In others, new or unforeseen problems arose only after the emergence of new technologies.[12]

Finally, and most broadly, the history of changing representations of the fossil record reveals a corresponding shift in the way paleontologists have understood the nature of their data and the kinds of knowledge derivable from it: a shift from imagining the record as a visual collection of objects that can be organized in physical or virtual space (the specimen cabinet or the illustrated catalog), to an abstract, randomly-accessible collection of data points (a database) existing in information "cyberspace" (for example, the magnetic storage drive). As I have argued, this shift began long before the emergence of computerized databases, but was perhaps fully realized during the age of computers. In their study of scientific atlases, Daston and Galison remark that such books function in part to "create one sliver of the world anew in images" (Daston and Galison, 2007, p. 27); similarly, paleontological compendia and databases re-create the past, and as the collection moved from a record of things to one of information, the past was re-created as data.

This is clearly part of a much broader phenomenon in the history of the natural sciences. One observation that is emerging is that while data-

[12] Peter Galison has described a similar relationship between technology and epistemic values in the development of microphysics. See Galison, 1997, Chap. 1.

driven research – or the "data deluge" – may be especially characteristic of twenty-first century science, its roots extend back into the eighteenth and nineteenth century practices of natural historians. This paper has attempted to demonstrate this through an analysis of paleontology, but, as Strasser has noted, "natural history has been 'data-driven' for many centuries" (Strasser, 2012, p. 87). Each era has had its own "information overload" to deal with, and while practices of collection, description, ordering, curating, and rendering objects into data may have changed over time, the basic values and requirements of data have remained remarkably constant. The specific practices that change, however, are not unimportant: in general, practitioners of the natural sciences during the eighteenth and nineteenth centuries had a much closer relationship with the physical objects that were the basis of their data than do many scientists involved in data analysis today. Strasser cites the "omnipresence of statistical methods" whereby data became "grist for statistical mills" as an "important novelty" in late-twentieth century science (Strasser, 2012, p. 87). While I would point out that, in paleontology, Lyell, Bronn, and others were quite preoccupied with processing their data numerically, I agree that statistics has taken on a new significance in the past 50 years. Many of the paleontologists who now specialize in quantitative analysis of the fossil record do, in fact, often work at significant remove from the field practices of their colleagues, and this has created tension within paleontology between the "traditionalists" and the "data jockeys."

Broadly, what these themes point to is the need for a richer history of the conceptualization and practices around data in the natural sciences of the past 300 years, and of the relationship between genres of information representation as diverse as the catalog and the database. From a technical, and perhaps also an epistemic, perspective, there are clear departures in the shift from early-nineteenth to early twenty-first century data practices, but also significant continuities. Strasser has posed the question of whether, in the biological and biomedical sciences "databases represent 'homologues' (the result of historical continuity) or 'analogues' (the result of a functional convergence) of the natural history museums and other naturalist collections" (Strasser, 2011, p. 94). My study of the development of data practices in paleontology strongly suggests a homologous relationship between nineteenth century fossil compendia and the twenty-first century database. If this is the case, then the history of the development of data practices in paleontology can shed light on both the emergence of the database as a central feature in modern natural science, and on the relationship the modern biological and biomedical sciences have to the practices of ‘‘collecting, describing, naming, comparing, and organizing natural objects’’ long associated with the traditions of natural history (Strasser, 2011, p. 62).

Acknowledgments

The first draft of this paper was written during a visit to the Max Planck Institute for the History of Science in Berlin during Fall, 2011. I would like to thank the members of the Department II colloquium there, and especially Lorraine Daston, for their helpful feedback and generous support. I would also like to thank Bruno J. Strasser for immensely insightful commentary on subsequent drafts of the paper, and Martin Rudwick, Paul Brinkman and Scott Lidgard for helpful suggestions about sources and problems along the way.

References

Archibald, J. David. 2009. ‘‘Edward Hitchcock’s Pre-Darwinian (1840) ‘Tree of Life’.’’ *Journal of the History of Biology* 42: 561–592.

Benton, M. J., and Emerson, B. C. 2007. ‘‘How Did Life Become So Diverse? The Dynamics of Diversification According to the Fossil Record and Molecular Phylogenetics.’’ *Palaeontology* 50: 23–40.

Blair, Ann. 2010. *Too Much to Know: Managing Scholarly Information Before the Modern Age*. New Haven: Yale University Press.

Bowker, Geoffrey C. 2005. *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Brongniart, Adolphe. 1828a. *Histoire des végétaux fossiles, ou Recherches botaniques et géologiques sur les végétaux renfermés dans les diverses couches du globe, par M. Adolphe Brongniart*. Paris: G. Dufour et E. d’Ocagne.

——1828b. *Prodrome d’une histoire des végétaux fossilles*. Paris: F.G. Levrault.

Bronn, H. G. 1831. *Italiens Tertiär-Gebilde und deren organische Einschlüsse: vier Abhandlungen, mit einer Steindrucktafel*. Heidelberg: K. Groos.

——1835. *Lethæa Geognostica, oder Abbildungen und Beschreibungen der für die Gebirgs-Formationen bezeichnendsten Versteinerungen*, 1st ed. Stuttgart: E. Schweizerbart.

——1841. *Handbuch einer Geschichte der Natur*. Stuttgart: E. Schweizerbart’sche Verlagshandlung.

——1848. *Index palaeontologicus, oder, Übersicht der bis jetzt bekannten fossilen Organismen*. Stuttgart: E. Schweizerbart.

——1858. *Untersuchungen über die entwicklungsgesetze der organischen welt während der bildungs-zeit unserer erdoberfläche*. Stuttgart: E. Schweizerbart.

Browne, Janet. 1983. *The Secular Ark: Studies in the History of Biogeography*. New Haven: Yale University Press.

Cutbill, J. L., and Funnell, B. M. 1967. "Numerical Analysis of the Fossil Record." W. B. Harland, C. H. Holland, and M. R. House (eds.), *The Fossil Record*. London: Geological Society, pp. 791–820.

Cuvier, Georges, and Jameson, Robert. 1813. *Essay on the Theory of the Earth. Translated*… by Robert Kerr… With Mineralogical Notes, and an Account of Cuvier's Geological Discoveries, by Professor Jameson. Edinburgh: Blackwood.

d'Orbigny, Alcide. 1840. *Paléontologie francaise description zoologique et géologique de tous les animaux mollusques et rayonnés fossiles de France Terrains cretaces*. Paris: l'Auteur.

Darwin, Charles. 1859. *On the Origin of Species*. London: J. Murray.

Daston, Lorraine, and Galison, Peter. 2007. *Objectivity*, 2 printed edn. New York, NY: Zone Books.

Davidson, Jane P. 2008. *A History of Paleontology Illustration*. Bloomington, IN: Indiana University Press.

Deshayes, Gérard P. 1824. *Description des Coquilles Fossiles des environs de Paris. 1, Conchifères*. Paris: Auteur u.a.

Deshayes, Gérard Paul. 1860. *Description des animaux sans vertèbres découverts dans le bassin de Paris 1, Atlas*. Paris: Baillière.

Foote, Michael. 1999. "J. John Sepkoski, Jr. (1948–1999)." *Acta Palaeontologica Polonica* 44(2): 235–236.

Galison, Peter. 1997. *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.

Giebel, Christian Gottfried Andreas. 1847. *Fauna der Vorwelt mit steter Berücksichtigung der lebenden Thiere Bd. 1. Wirbelthiere. Abth. 1, Die Säugethiere der Vorwelt mit steter Berücksichtigung der lebenden Säugethiere*. Leipzig: Brockhaus.

Gliboff, Sander. 2007. "H.G. Bronn and the History of Nature." *Journal of the History of Biology* 40(2): 259–294.

——2008. *H.G. Bronn, Ernst Haeckel, and the Origins of German Darwinism: A Study in Translation and Transformation*. Cambridge, MA: MIT Press.

Harland, W. B., Holland, C. H., and House, M. R., Geological Society of London, and Palaeontological Association (eds.). 1967. *The Fossil Record*. London: Geological Society.

Herbert, Sandra. 2005. *Charles Darwin, Geologist*. Ithaca, NY: Cornell University Press.

Koerner, Lisbet. 1999. *Linnaeus: Nature and Nation*. Cambridge, MA: Harvard University Press.

Latour, Bruno. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Leonelli, Sabina. 2012. "Making Sense of Data-Driven Research in the Biological and Biomedical Sciences." *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 43(1): 1–3.

Lyell, Charles. 1830–1833. *Principles of Geology; Being an Attempt to Explain the Former Changes of the Earth's Surface, by Reference to Causes Now in Operation*. London: J. Murray.

Mantell, Gideon Algernon. 1822. *The Fossils of the South Downs; or Illustrations of the Geology of Sussex*. London: Lupton Relfe.

Morris, John. 1843. *A Catalogue of British Fossils. Comprising all the Genera and Species Hitherto Described; with References to Their Geological Distribution and to the Localities in Which they have been Found*. London: J. Van Voorst.

—— 1854. *A Catalogue of British Fossils: Comprising the Genera and Species Hitherto Described: With Reference to Their Geological Distribution and to the Locatities in Which they have been Found*, 2nd edn. Considerably Enlarged Edition. London.

Müller-Wille, Staffan, and Charmantier, Isabelle. 2012. "Natural History and Information Overload: The Case of Linnaeus." *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 4–15.

Ogilvie, Brian. 2003. "The Many Books of Nature: Renaissance Naturalists and Information Overload." *Journal of the History of Ideas* 64: 29–40.

Parkinson, James. 1822. *Outlines of Oryctology: An Introduction to the Study of Fossil Organic Remains*. London: Sherwood.

Patrik, Linda E. 2000. "Is There an Archaeological Record?" J. Thomas (ed.), *Interpretive Archaeology: A Reader*. London: Leicester University Press, pp. 118–144.

Phillips, John. 1835. *Illustrations of the Geology of Yorkshire; or, A Description of the Strata and Organic Remains of the Yorkshire Coast: Accompanied by a Geological Map, Sections, and Plates of the Fossil Plants and Animals*, 2nd ed. York: T. Wilson.

—— 1841. *Figures and Descriptions of the Palaeozoic Fossils of Cornwall, Devon, and West Somerset; Observed in the Course of the Ordnance Geological Survey of that District*. London: Longman.

——1860. *Life on the Earth; Its Origin and Succession*. London: Macmillan and Co.

Pickstone, John V. 2001. *Ways of Knowing: A New History of Science, Technology, and Medicine*. Chicago: University of Chicago Press.

Porter, Theodore M. 1986. *The Rise of Statistical Thinking, 1820–1900*. Princeton, NJ: Princeton University Press.

Rappaport, Rhoda. 1997. *When Geologists were Historians, 1665–1750*. Ithaca, NY: Cornell University Press.

Raup, David M. 1986. *The Nemesis Affair: A Story of the Death of Dinosaurs and the Ways of Science*. New York, NY: W. W. Norton & Co.

Raup, David M., and Sepkoski, J. John Jr. 1984. "Periodicity of Extinctions in the Geologic Past." *Proceedings of the National Academy of Sciences of the United States of America* 81(3): 801–805.

Rudwick, Martin J. S. 1975. "The Visual Languages of Geology in the Early 19th Century." D. A. Bassett (ed.), *Charles Lyell Centenary Symposium; Programme and Abstracts*. International Union of Geological Sciences, International Committee on the History of Geological Sciences, 34.

——1976. *The Meaning of Fossils; Episodes in the History of Palaeontology*, 2nd ed. New York, NY: Neale Watson Academic Publishers, Science History Publishers.

—— 1978. "Charles Lyell's Dream of a Statistical Palaeontology." *Palaeontology* 21(Part 2): 225–244.

—— 2000. "Georges Cuvier's Paper Museum of Fossil Bones." *Archives of Natural History* 27(1): 51–68.

—— 2005. *Bursting the Limits of Time: The Reconstruction of Geohistory in the Age of Revolution*. Chicago: University of Chicago Press.

Sepkoski, J. John Jr. 1981. "A Factor Analytic Description of the Phanerozoic Marine Fossil Record." *Paleobiology* 7: 36–53.

—— 1982. *A Compendium of Fossil Marine Families*. Milwaukee: Milwaukee Public Museum.

—— 1984. “A Kinetic Model of Phanerozoic Taxonomic Diversity. III. Post-Paleozoic Families and Mass Extinctions.” *Paleobiology* 10(2): 246–267.

—— 1992. *A Compendium of Fossil Marine Animal Families*. Vol. 83, *Contributions in Biology and Geology*. Milwaukee, WI: Milwaukee Public Museum.

—— 1994. “What I Did with My Research Career; or How Research on Biodiversity Yielded Data on Extinction.” W. Glen (ed.), *The Mass-Extinction Debates; How Science Works in a Crisis*. Stanford, CA: Stanford University Press, pp. 132–144.

Sepkoski, David. 2012. *Rereading the Fossil Record: The Growth of Paleobiology as an Evolutionary Discipline*. Chicago: University of Chicago Press.

Sepkoski, David, and Raup, David M. 2009. “An Interview with David M. Raup.” D. Sepkoski and M. Ruse (eds.), *The Paleobiological Revolution: Essays on the Growth of Modern Paleontology*. Chicago: University of Chicago Press, pp. 459–470.

Sepkoski, David, and Ruse, Michael. (eds.). 2009. *The Paleobiological Revolution: Essays on the Growth of Modern Paleontology*. Chicago: University of Chicago Press.

Sepkoski, J. John Jr., Jablonski, David, and Foote, Michael. 2002. *A Compendium of Fossil Marine Animal Genera, Bulletins of American Paleontology. 363; 2002*. Ithaca, NY: Paleontological Research Institution, p. 560.

Smith, William. 1816. *Strata Identified by Organized Fossils*. London: W. Arding.

Strasser, Bruno J. 2011. “The Experimenter’s Museum: GenBank, Natural History, and the Moral Economies of Biomedicine.” *Isis* 102(1): 60–96.

—— 2012. “Data-Driven Sciences: From Wonder Cabinets to Electronic Databases.” *Studies in the History and Philosophy of Biological and Biomedical Sciences* 43: 85–87.

Whewell, William. 1837. *History of the Inductive Sciences: From the Earliest Times to the Present*. London: John W. Parker.

Woodward, John. 1729. *An Attempt Towards a Natural History of the Fossils of England: In a Catalogue of the English Fossils in the Collection of J. Woodward, M.D.* London: Fayram.

Woodward, Samuel. 1830. *A Synoptical Table of British Organic Remains; in Which, all the Edited British Fossils are Systematically and Stratagraphically Arranged, in Accordance with the Views of the Geologists of the Present Day*…. London: Longman.

Zittel, Karl A von. 1901. *History of Geology and Palaeontology to the End of the Nineteenth Century*. Translated by M. M. Ogilvie-Gordon. London: Walter Scott.