

# When Are Fossils Data?

Aja Watkins\*

---

Existing accounts of data are unclear about whether the epistemic role objects play makes them data, or whether data have to be produced by human interaction with the world – these two features can come apart. I illustrate this ambiguity using the case of fossil data, which have rich histories and undergo many processes before they are encountered by humans. I then outline several philosophical positions that would resolve the ambiguity moving forward, and elaborate on my preferred option.

---

## Keywords


data • fossils • fossil data • paleontology

## 1 Introduction

Philosophers of science have recently become increasingly interested in characterizing data and analyzing data-related scientific practices. For example, according to Sabina Leonelli's (2015) influential "relational" view of data, data are defined as follows: an object is a piece of data when "(1) it is treated as potential evidence for one or more claims about phenomena, and (2) it is possible to circulate it among individuals" (817). Leonelli's view calls attention to the material nature of data, and also helps to explain the now-received view among philosophers that no data are "given"; data are made by researchers to serve particular roles in inquiry.

Alisa Bokulich and Wendy Parker (2021) present an alternative view of data, which they call the "pragmatic-representational" view. According to them, "data are representations that are the product of a process of inquiry, and they should be evaluated in terms of their adequacy or fitness for particular purposes" (1). Their main disagreement with Leonelli is about whether data represent; according to Leonelli (2019), only data *models* – among other kinds of models – represent (she prefers to distinguish between the evidential role of data and the representational role of models).

---

\*Department of Philosophy, University of Wisconsin-Madison, Madison, WI, USA, aja.watkins@wisc.edu  
 <https://orcid.org/0000-0003-0171-8738>

Received 4 May 2023; Revised 23 January 2024; Accepted 26 March 2024  
doi:10.3998/ptpbio.5521

In this paper, I will point out an ambiguity pervasive in both of these existing philosophical accounts of data, as well as in other discussions thereof. The ambiguity concerns when, exactly, an object becomes a piece of data. In the examples relied upon by Leonelli, as well as other examples utilized by Bokulich and Parker, two moments tend to roughly coincide: the moment when an object is taken to potentially serve as evidence (or to serve as a representation) by investigators and the moment at which the object is constructed, measured, or otherwise encountered by those same researchers. I will refer to these throughout as the moment when an object acquires a certain *epistemic function* and the moment when an object is produced via a certain *human-world interaction*, while remaining uncommitted to exactly which epistemic function and exactly which human-world interaction are the important ones. As I will illustrate using the case of fossil data, the moment at which an object acquires the right kind of epistemic function and the moment at which an object is produced via the right kind of human-world interaction (whatever these may be) need not co-occur, and, indeed, vast amounts of time can elapse between these moments for certain objects.

The same ambiguity also arises in characterizations of data journeys, which, according to Leonelli (2020), “can be broadly defined as designating the *movement of data from their production site to many other sites in which they are processed, mobilized and repurposed*” (9; emphasis original). The ambiguity here is what, exactly, counts as the data’s “production site.” Likewise, as I’ll argue, questions about what counts as metadata and what counts as “data-time” versus “phenomena-time” (terminology from Leonelli 2018) also relate to the question of when an object becomes data.

In other words, current work in philosophy of data tends to agree that data is a *relational* category – something isn’t data or not full-stop, but only as the result of existing in the right kind of relationship – but disagree on what, exactly, constitutes the “right kind” of relationship. (Is it an evidential relationship? representational?) However, amidst these disagreements lingers a background question about what, even, the appropriate *relata* of that relationship are. Is an object data in relation to a particular question or claim (as an account of data that focuses on the epistemic function of data would suggest)? Or is an object data in relation to a particular researcher or set of researchers, their activities, and their pursuits? This paper argues that present philosophical accounts of data have not answered this question, and uses the case of fossil data to both try to motivate the question and survey the possible answers.

The structure of the paper is as follows. First, in section 2, I’ll give a brief explanation of how fossil data are generated, including some commentary on the various times at which we might think the fossils become data and why it’s not obvious where to draw the line. In section 3, I’ll present the ambiguity in more detail, contextualized within contemporary philosophical accounts. Finally, in section 4, I’ll present several options for how the ambiguity might be resolved moving forward, including some (defeasible) reasons for which route I think is the most plausible.

## 2 Fossil Data

Philosophers of data, perhaps surprisingly, do not dispute that prepared fossils – those ready for use by paleontologists or in museum displays – are data. Bokulich (2021) says this outright: “fossil rocks ... can be thought of as a physical data model. The fossils in this context are taken as a representation of past life on Earth” (17; recall Bokulich and Parker’s representational view of data). Leonelli’s relational view of data also implies that fossil specimens themselves count as data. (They are treated as potential evidence and it is possible to circulate them, so they qualify according to her definition; cf. Wylie [2019].) In general, Leonelli is quite explicit that physical

objects such as specimens or model organisms count as data on her view (e.g., 2015, 817), and, indeed, she takes data in the historical sciences as inspiration for her account of data-time and phenomena-time in Leonelli (2018), to which we will return in section 3.

However, fossils have a long history, and fossil data are shaped by many processes that occur before researchers ever get involved. (For more information, see classic work done by Shipman [1981] and an excellent contemporary discussion by Holland [2016].) I will argue below that structural similarities between the processes that shape fossil data before those fossils are discovered and the processes that later shape fossil data as the result of human interaction with the fossil specimens makes it difficult to say when, exactly, a fossil becomes a piece of data.

Processes that shape fossil data but which occur before humans ever encounter the fossil specimen include:

- **Death and Burial.** How an organism dies affects the likelihood that it and its various features will be preserved in fossilized form. The nature of an organism's death can thus affect its usefulness to scientists for reconstructing the nature of past organisms or ecosystems. If death is the result of predation, for instance, remains may be scattered about or digested. Additionally, organisms have to be buried in the right kinds of sedimentary environment (e.g., a lacustrine environment with high sedimentation rates) to be fossilized and preserved.<sup>1</sup> Finally, whether an organism is buried immediately after death or after decomposition affects the state in which it is fossilized, and, consequently, the state in which it may later be discovered.
- **Mineralization.** Fossils do not have the same material constitution as organisms' remains. During diagenesis – conversion of sediment layers into sedimentary rock – fossil remains become mineralized.<sup>2</sup> One way this can happen is through permineralization, during which minerals fill pores in organic remains, replacing organic matter as it disintegrates. Entirely permineralized remains are “petrified” (turned to stone). Mineralization can cause plastic deformation of remains, changing their shape and dimensions.
- **Preservation.** Depositional/erosional environments are differentially conducive to preservation, with some fossils in “doomed sediments” that are highly likely to erode (or be subducted), a likelihood which varies systematically by taxa, depositional environment, sedimentary basin type, and time period. Erosion leads to the “pull of the recent,” whereby more fossil evidence is available for closer periods of time (leading to biased biodiversity estimates; Raup 1972). Furthermore, the “preservation potential” of different kinds of bones is different; for example, teeth are especially likely to fossilize due to their density. Additionally, Lagerstätten (deposits with exceptionally high preservation) bias the fossil record of various taxa (e.g., pterosaurs; Dean, Mannion, and Butler 2016).
- **Relocation.** Movement of fossils post-deposition, caused by Earth's shifting crust, can impede reconstructions of location and order of deposition, although today many data correction tools exist to account for this movement (with some uncertainty). Relocation of fossils also affects the likelihood they will be found by humans via an “exposure bias.”

---

1. Ideal preservation environments may differ systematically between vertebrate and invertebrate taxa, or between plants and animals, both as a function of where these organisms are more likely to live and of what kinds of materials they are composed.

2. This primarily applies to *bones*. For many marine organisms with shells, the original shell material can be preserved for millions of years. Other fossils, such as of footprints, are just impressions or molds. (These are called “trace fossils”; see Finkelman [2019] for discussion.)

Fossils are often found at rock outcrops, exposed (more accessible) areas of rock. However, this requires that the depositional area in which the fossil was emplaced is now an erosional area; this can occur through literal movement of the strata or through a change in conditions surrounding strata that have not moved. Whether a fossil is located in an accessible rock outcrop is a more-or-less random selection procedure that creates a sampling bias in the fossil record as we know it, and affects our subsequent efforts to track biodiversity over time. Correcting for this sampling bias has preoccupied paleontologists for some time (e.g., Raup 1976; Alroy et al. 2001; Smith 2001; Peters and Foote 2001; Peters 2005; Smith and McGowan 2005; Lloyd et al. 2012; Lloyd, Young, and Smith 2012).

Of course, fossils are also processed and altered by humans starting when they are discovered and excavated. (For excellent history and sociology of science work on these processes, see Rieppel [2019] and Wylie [2021].) These processes, like those that happen before humans interact with the fossil, shape which fossil data are available to researchers and how they can be used. Some of these processes include:

- **Discovery.** To be studied, fossils need to be found and, most likely, to also be extracted from rock formations in which they were preserved. While fossil discovery can be serendipitous, it is nevertheless theory- and value-laden. For starters, fossil “discoveries” in the United States have historically only been made by white (male) colonizers/settlers, without recognizing that many Indigenous peoples were previously well-aware of the presence of fossils (Mayor 2013). Additionally, fossil discoveries often have occurred alongside prospecting missions for other natural resources, such as coal; this has provided a reliable way for fossil hunters to finance their searches and fossilized remains have sometimes been uncovered accidentally in the course of mining activities (Rieppel 2019, chap. 1). Consequently, fossil discovery locations are not random, and the locus of major fossil discoveries (at least those recognized by Western science) roughly tracks changes in geopolitical power and scientific prowess: Europe (early nineteenth century), North America (mid-nineteenth to mid-twentieth century), and now China. Furthermore, even for a given fossil being *uncovered*, there is often debate about whether a new fossil counts as a *discovery* (e.g., of a new taxon), related to age-old taxonomic debates about “splitting” versus “lumping.”
- **Excavation and Transportation.** Excavation and transportation of fossils is affected by background assumptions and material constraints. When excavators remove fossils and the surrounding rock matrix, they make decisions about which tools to use, which fossils are worth removing, and which parts of the rock matrix (the material in which the fossil is embedded) to include with the fossil upon extraction. Lukas Rieppel (2019) shows how these practices have historically been informed by extractive mining, not only because fossils and mined resources were often discovered together but also because fossil excavators themselves often had a mining background. Excavators also make judgments about which fossil samples are most likely to be productive, scientifically or financially. Excavation may also alter crucial characteristics of the fossil assemblage. Excavators have developed practices to document the position of the fossilized remains in the rock bed, metadata that is later used to help reassemble the organism or fill in gaps. Likewise, decisions about how to transport fossils also require judgment, such as about which materials to use to secure the fossil, what means of transportation are appropriate, and how to label the extracted fossils, keeping in mind the fragility of the fossil and its likely future uses.

There are also patterns concerning where fossils are transported *from* (the field locations where they are discovered) and where they are transported *to* (museums and labs, likely located in affluent regions); Raja et al. (2022), for example, criticize these practices for their historical relationship to colonialism and exploitation and trace their likely epistemic consequences (see also: Liboiron 2021; Monarrez et al. 2021; Cisneros et al. 2022; Wylie 2024). Innovations in transportation (cross-continental railroads, air travel) have made these trips more or less practical in different contexts throughout history. Furthermore, technological innovations such as casting or scanning fossils *in situ* may remove the need to excavate and transport them.

- **Preparation and Conservation.** Caitlin Wylie (2021) performed an ethnography of vertebrate fossil preparators, the technicians responsible for turning fossils into specimens usable by paleontologists or museum curators. Her work highlights that although fossil preparators do not literally make fossil specimens from scratch, they nevertheless are faced with creative decisions, including the best ways of “separating useless background (such as matrix) from informative specimen” and “decisions about what an object should look like, and what it should be capable of” (Wylie 2021, 4). For example, preparators decide how much rock matrix to scrape away: a desire to leave only fossilized remains counts in favor of removing more matrix, but risks damaging the fossil. Preparators also decide whether to use removable or permanent glue, a decision which changes the chemical constitution of the fossil and constrains its future uses. Preparators, conservators, and scientists are divided about whether to protect possible future research needs or prioritize current research. Preparators also decide whether to use new techniques or rely on old ones; for example, recently, preparators have had the option of using CT scanning to prepare a “digital fossil,” but they have often chosen to rely on century-old techniques instead (Wylie 2021, chap. 3). CT scans are preferred when fossils are too fragile to prepare, whereas traditional preparation is better for examining fossilized material itself. Wylie (2019) thus argues that prepared fossils used by scientists are underdetermined by unprepared fossils delivered to the preparator, and recommends the use of additional metadata on fossil preparation techniques used in order to allow future researchers to reconstruct fossils’ history (see also Wylie 2016).
- **Assembly.** Especially for museum exhibits, fossils have to be assembled as “models” of the organisms they represent (Nyhart 2004). Many decisions have to be made to assemble a fossil to (purportedly) resemble the original organism. First, fossils are almost always incomplete; if a complete assembled fossil is desirable, gaps need to be filled. Fossil preparators sculpt missing pieces based on anatomical expectations or use bones from other, similar remains. Historically, there has been some debate over whether to make the inauthentic pieces of an assembled fossil more explicit or more discrete (Rieppel 2019). At this stage, some fossils may be replicated using plaster casts, which enables specimens’ circulation to other museums or laboratories. Second, what kind of stand or other support structure to use needs to be decided upon, depending on the specimen’s fragility. Finally, assemblers have to decide how, exactly, to orient the pieces in order to make the assembly. This decision requires anatomical and biomechanical considerations, as well as rhetorical considerations such as how imposing or docile to make the assembly seem.<sup>3</sup>

3. Some famous controversies regarding fossil assembly include Edward Drinker Cope mounting the vertebral column of *Elasmosaurus platyrus* backwards (with its skull at the tip of its tail; ca. 1868), and the debate about whether sauropod dinosaurs’ legs should be below the torso, like elephants, or to the side, like lizards (Turner 2007, 89–91).

- **Artistic Depiction.** With some exceptions, only hard parts of organisms are preserved through fossilization. Mineralized skeletal remains are adequate for research purposes related to morphology, behavior (e.g., ambulation), and diet (via tooth analysis). However, other purposes require that unfossilized parts of an organism (i.e., soft tissues) be reconstructed. This need can be filled by artistic depiction of organisms from the deep past, including paintings, sculptures, and animations (for historical analysis of paleoart, see Mitchell 1998; Noble 2016; Lescaze 2017). Artists work in connection with paleontologists to fill in details regarding musculature, skin texture, coloration, and ecological setting.<sup>4</sup> Scientists' and artists' preconceptions tend to (literally) color their depictions of past organisms. For example, an association between dinosaurs and reptiles long led to depictions of dinosaurs as having scaly skin; recognition that dinosaurs are more closely related to modern-day birds, as well as several discoveries of fossilized impressions of feathers, has now led to depictions of many dinosaurs as feathered.<sup>5</sup>

All of these stages a fossil specimen goes through are depicted in figure 1.

Importantly, there are structural similarities between some of the processes that occur pre-excavation and those that occur post-excavation. These processes can be categorized as follows: First, some processes both pre- and post-discovery can *filter* the fossil data. Data filtering involves deciding which parts of the specimen make it to the next step. For example, decomposition removes some parts of the organism (e.g., soft tissue), deposition into a sedimentary environment conducive to fossilization enables addition of the specimen into the rock record, erosion might remove all or part of a specimen from that record, fossil preparation removes some parts of the rock matrix, fossil assembly might involve adding synthetic or copied parts of a fossil to fill in “gaps,” and artistic depiction can involve adding phenotypic or ecosystemic information. These various processes of inserting or deleting information and substance shape the eventual fossil data or data models that are used by paleontologists or consumed by the public.

Second, some pre- and post-discovery processes change the *material constitution* of the fossil specimens. For example, mineralization converts organic remains into fossil rocks, and some steps performed during fossil preparation and assembly physically alter the material of the specimen (e.g., by adding glue to repair a broken specimen). Thus the material features of the fossil specimens can be altered both before and after they are discovered and investigated by humans, and both of these processes shape the eventual use of the fossil as data.

Finally, both pre- and post-discovery processes include those that *relocate* the fossil specimens. Various geological processes can relocate organismal remains (before or after fossilization), and much of human-generated fossil processing involves moving specimens from one location to another (e.g., from the extraction site to the preparation site). Again, these processes affect the resulting fossil data in similar ways, indicating that there is not a difference in kind between the processes that affect fossil data before those fossils are discovered by humans and those that affect fossil data afterwards.

Of course, the *causes* of pre- and post-discovery processes are very different: the first set are caused by geologic and taphonomic processes, whereas the second set are caused by intentional action. But the *effects* are similar in kind. Indeed, one important piece of information about

4. Benton (2019) describes this vaguely as a “conversation” between the paleontologist and the artist, using “clues” from the organism’s skeleton, highlighting the difficulty of articulating exactly how the decisions that go into an artistic depiction are made. Recently, some automatic digital tools have become available to assist in soft-tissue reconstruction; see Lautenschlager (2016).

5. Investigations into the color of dinosaurs has received some philosophical attention (Turner 2007, 2016, 2019; Jeffares 2010); we now have techniques to study melanosomes in fossilized dinosaur parts to reconstruct their most likely coloration (e.g., Vinther et al. 2008).

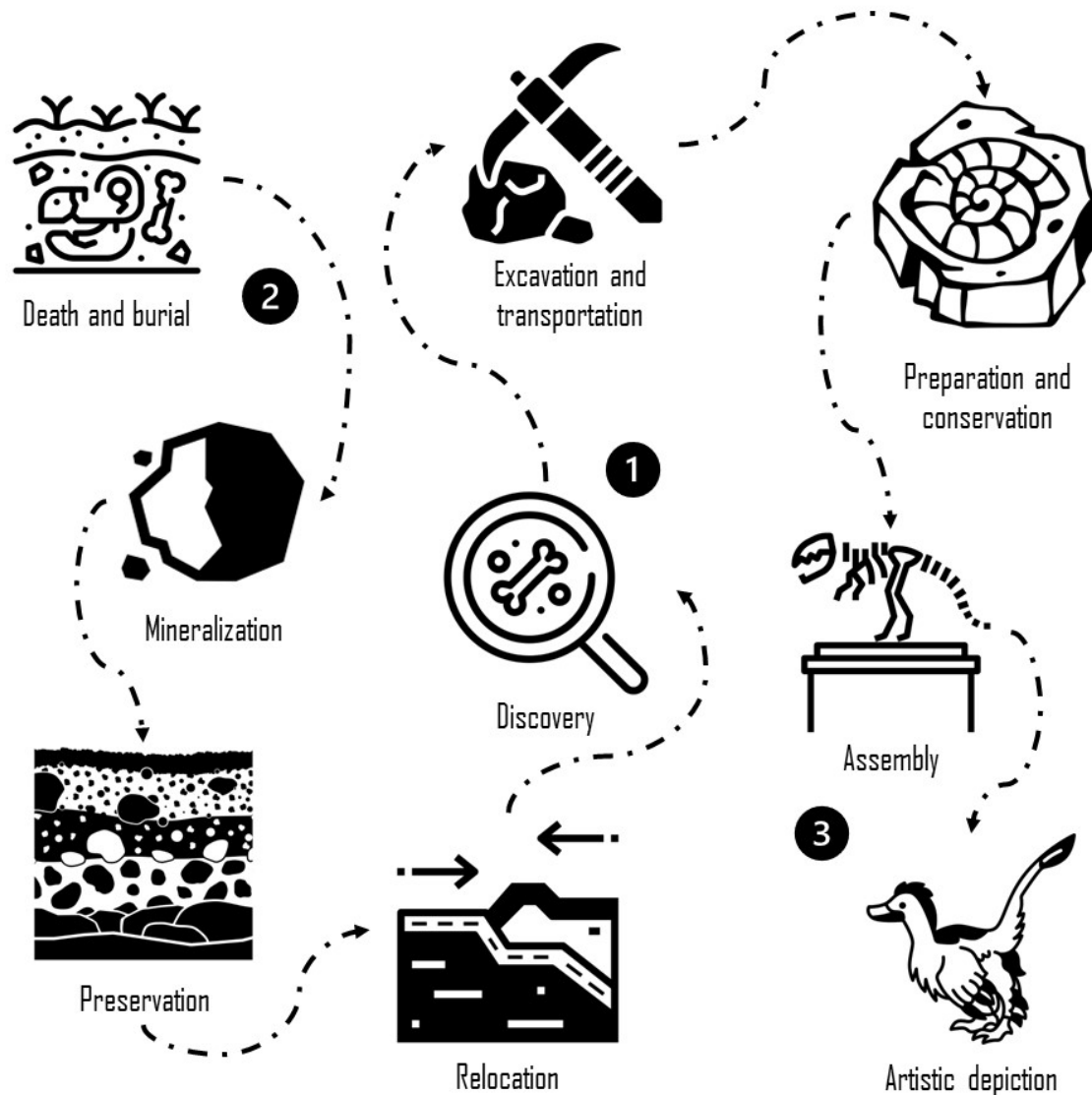


Figure 1: The various processes that affect and produce fossil data. Processes both before and after fossil discovery can filter the data (e.g., decomposition, deposition, erosion, preparation, assembly), alter its material constitution (e.g., mineralization, preparation), or relocate it (relocation can happen as a result of geologic or anthropogenic activities). Three places are labeled that might be the moment at which a fossil becomes a piece of data: (1) the moment of discovery, (2) the time at which the specimen becomes a fossil at all, or (3) the time at which the fossil specimen actually starts being used as data by paleontologists.

paleontological practice is that paleontologists refer to information about how a fossil specimen is “processed” geologically and taphonomically as “metadata.” I’ll return in section 3 to whether that is a fair characterization, and only note for now that paleontologists, at least, do not think that information about how their specimens are affected pre- or post-discovery needs to be characterized differently.

Although I have divided up the processes listed above into those that occur pre-discovery and post-discovery, there are no pre-theoretic reasons to say that the moment of discovery is a salient moment for determining when the fossil specimens become data. Certainly we could just apply existing philosophical definitions of data to answer this question (and I will return to these definitions in detail in section 3), but the exact question under consideration in this paper is whether those definitions are adequate, and, more specifically, whether they adequately address the case of fossil data.

I think that, if we put existing philosophical accounts of data to one side, there are roughly three salient places at which a fossil specimen might become a piece of fossil data. First, a fossil specimen might become a piece of fossil data at the time when it is discovered or collected; I think that this is the most intuitive place to draw the line. (This option is labeled with the number “1” in figure 1.) However, taking this option has the major downside of failing to recognize the structural similarities between pre- and post-discovery fossil specimen processing outlined above, each of which can affect eventual fossil data in similar ways. Second, and perhaps in response to noticing the aforementioned structural similarities in specimen processing, one might decide that a fossil specimen is data all along – from the moment it becomes a fossil at all. (This option is labeled with the number “2” in figure 1.) This option recognizes that even undiscovered fossils play a potentially significant role in inquiry (for example, they serve as potential evidence). However, a counter-intuitive consequence of this second option is that some objects which may *never* be encountered by human researchers, such as some fossils that are never discovered or are destroyed before they could have been discovered, would still count as data. Third, some might argue that, actually, fossils aren’t data until paleontologists or other scientists actually use them as such, which wouldn’t happen until the very final stages of the specimens’ processing. (This option is labeled with the number “3” in figure 1.)

In the following section, I will argue that existing philosophical accounts of data (and data journeys, metadata, and data-time) do not provide sufficient resources to decide when, exactly, fossils become data. Furthermore, the fossil case helps to illustrate that the ambiguity generalizes: existing accounts of data are ambiguous about whether an object becomes data when it serves a particular *epistemic function* (like serving an evidential or representational role) or whether it becomes data when it is the product of a certain kind of *interaction* between investigators and the world.

### 3 An Ambiguity in Philosophy of Data

Existing philosophical accounts of data do not clear up where, exactly, a fossil specimen becomes data. That is because there is an unresolved ambiguity concerning when any object, including a fossil, would become a piece of data. In this section, I illustrate the ambiguity in three ways. In section 4, I review options for resolving the ambiguity, and present considerations in favor of or against each.

First, the ambiguity arises in accounts of what data *are*. Leonelli (2015) has offered a very influential account of data, which she uses to reorient philosophers of data to data practices in science as opposed to merely the age-old conundrum of how data can serve as evidence if all observations are theory-laden. In Leonelli’s account, she provides two jointly sufficient require-



ments for anything to count as a piece of data: “(1) it is treated as potential evidence for one or more claims about phenomena, and (2) it is possible to circulate it among individuals” (Leonelli 2015, 817). Leonelli says that she wants to “give up altogether on a definition of data based on the degree to which they are manipulated and focus instead on the relation between researchers’ perceptions of what counts as data and the stages and contexts of investigation in which such perceptions emerge” (Leonelli 2015, 817). Similarly, Leonelli (2019) says “whether a set of objects functions as data or models does not depend on ... the degree of human intervention involved in generating them, but rather on their distinctive roles towards identifying and characterizing the targets of investigation” (Leonelli 2019, 2). According to these passages, it seems that what is important about data is their evidential role, such that something becomes data when it is recognized as potentially playing a certain role in scientific investigation (i.e., serving as potential evidence in a given research context). However, elsewhere Leonelli *does* seem to assign importance to human action as producing data. For example, Leonelli (2016) says data are “defined by the evidential value ascribed to them at specific moments of inquiry” (70), but later “propose[s] to define data as any *product of research activities* ... that is collected, stored, and disseminated in order to be used as evidence for knowledge claims” (Leonelli 2016, 77; emphasis added). Likewise: “*Data are the results of interactions between researchers and the world*, which are construed and processed to function as usable evidence for claims about phenomena” (Leonelli 2019, 23; emphasis added). The ambiguity, then, is whether or to what extent human actions such as measurement or data collection are characteristic of the moment when something becomes data, or whether taking something to serve as potential evidence – without necessarily physically interacting with it in any way – counts as the kind of “interaction” that produces data. Of course, measuring a phenomenon or collecting data might be one way in which data are produced; these are cases in which the moments of human involvement and ascribed evidential value will coincide. But there may be *other ways* of treating an object as potentially evidential, in which case manipulation or measurement would not be necessary to make something into data.<sup>6</sup>

For example, in the fossil case focused on throughout this paper, only the post-discovery fossil specimens could reasonably be construed as a “product of research activities” or as “the results of interactions between researchers and the world.” However, pre-discovery fossils certainly have *potential* evidential value (that potential will be actualized if and when the fossil is discovered, extracted, prepared, and used in paleontology research). It is also, in a sense, *possible* to circulate never-before-seen-by-us fossils, as it is possible to move around all kinds of mid-sized objects. Thus, pre-discovery fossils seem to meet both of Leonelli’s criteria for being pieces of data, while failing to have her implicit or assumed characteristics of data, namely, that data are produced by us.<sup>7</sup>

6. An additional complication comes from the fact that taking on a particular epistemic role (like serving as potential evidence) isn’t as easily indexed to a moment in time as a human-world interaction is. Thanks to Caleb Hazelwood and Meghan Page for helping me to think about this point.

7. A useful distinction might be made between “potentially evidence” and “contingently evidence,” or, correspondingly, between “serving as potential evidence” and “potentially serving as evidence.” (Thanks to Annette Zimmermann for suggesting this.) If an object “serves as potential evidence,” that would involve treating the object as though it might, someday and in some context, serve as evidence; undiscovered fossils serve as potential evidence in this way. If an object “potentially serves as evidence,” that would involve treating the object as though it is ready to be deployed as evidence at any moment; undiscovered fossils do not potentially serve as evidence in this way. Leonelli herself does not distinguish between these two; technically, she says “treated as potential evidence,” which sounds more like the former than the latter, in which case undiscovered fossils would meet the first criterion. More work is needed to flesh out the details of what “potentially” means in Leonelli’s account, both in terms of “potential evidence” and in terms of “possible to circulate.”

The same ambiguity arises in Bokulich and Parker's (2021) pragmatic-representational account of data. On the one hand, Bokulich and Parker say that data (and data models) are representational, i.e., "*taken to be about* one or more aspects of the world" (Bokulich and Parker 2021, 7; emphasis original). This characterizes data in terms of their epistemic function (in their case, the emphasis is on representational rather than evidential function). On the other hand, though, Bokulich and Parker embrace "what should be uncontroversial: that data are the product of an *interaction* between a measuring device (or observer) and the world" (9; emphasis original). Again, then, it is unclear whether data are primarily characterized by their epistemic use or by their production by humans, or whether, perhaps, both are required.

Again, epistemic function and production by humans might correspond, especially in cases where the act of observing or measuring something *is* what enables it to serve as evidence. The cases explored by Leonelli (2015) and Bokulich and Parker (2021) are all like this, while the paleontology case described herein is not. As the paleontology case highlights, something can be potentially evidence without ever having come into contact with humans. Consider that Leonelli (2016) allows some organisms to count as data, especially model organisms which, she argues, are literally made by researchers (see also Ankeny and Leonelli 2021). However, organisms in the "wild" may also serve as data according to the view that characterizes data by their epistemic function, a view on which what it is to be data is just to exist in the right relation to a claim or research question; for instance, these organisms (intuitively) serve as potential evidence in biology (and are portable). It is not clear, though, whether these organisms *qua* data are produced when they enter into the appropriate relationship to a claim or research question (e.g., an evidential relationship), or whether *qua* data they are produced when they enter into the appropriate relationship with researchers (e.g., are observed or measured by them). In general, the ambiguity is epitomized by objects that are known to be potential sources of evidence in relation to a given research context, but haven't yet been observed.

A second way to characterize the ambiguity arises in the literature on data journeys, where it is also unclear how much emphasis to put on human-world interaction rather than epistemic function (where epistemic function could be a potential evidential role, a representational role, etc.). Recall that Leonelli (2020) says "Data journeys can be broadly defined as designating the *movement of data from their production site to many other sites in which they are processed, mobilized and repurposed*" (Leonelli 2020, 9; emphasis original). Some data journeys involve movement of objects between physical locations, whereas others involve other transformations, such as between different file formats. Data journeys need to be reconstructed in order to appropriately interpret and use data. The need to do so raises the question of precisely when data journeys begin: where exactly is the data's "production site"? One obvious place to draw the line is at the moment of researcher-world interaction, e.g., measurement or observation or discovery. However, as is illustrated by the case of fossil specimens, some data have rich, important histories prior to the moment of interaction with researchers. Information on these pre-interaction histories are often included as metadata and accounted for in similar ways as other data journeys are accounted for when deciding how to use and interpret data. For instance, Leonelli (2019) calls all "related contextual information" metadata, including, for example, "plant provenance and growth conditions," which may include transformations before researchers get involved (Leonelli 2019, 11). Whether we characterize data based on epistemic function or production by scientists may change where we decide to delineate data journeys and other processes which affect data interpretation and use.

Nora Boyd (2018) also emphasizes the importance of metadata. She argues that scientific inquiry requires "enriched evidence," evidence supplemented with metadata, which Boyd characterizes as "auxiliary information about empirical results" (Boyd 2018, 410). Boyd says there

are two types of metadata: “provenance’ metadata (associated with the data collection stage of research) and ‘work-flow’ metadata (associated with the data-processing stage of research)” (Boyd 2018, 410). According to this classification, the important auxiliary information only encompasses what happens after investigators interact with their research subject. But why not think that auxiliary information about provenance extends further back in time? On an account of metadata that focused more on its epistemic function in scientific reasoning, all provenance-related information would be potentially important. Of course, one view is that when data journeys begin or when metadata becomes relevant is not necessarily at the same time that something becomes data – but both cases involve the need to determine whether the act of data collection is ontologically important or whether something serving the epistemic function of data, metadata, or data journeys is enough.

A third way to illustrate the same ambiguity has to do with Leonelli’s (2018) distinction between phenomena-time and data-time. Data-time is “the time at which data collection, dissemination and analysis occur,” whereas phenomena-time is “the time in which phenomena for which data serve as evidence operate” (Leonelli 2018, 741). Data-time begins at the moment of human interaction with the research subject. However, phenomena-time is characterized by the epistemic target of research. The moment at which human agents get involved and the moment which concludes the phenomena for which data serve as (potential) evidence may or may not coincide, such that there may be overlaps between data-time and phenomena-time (a period of data collection during which the relevant phenomena are still occurring) or a period of neither phenomena-time nor data-time (a period where the relevant phenomena have ceased but before data collection). Again, concrete objects like fossils help to illustrate the problem, because concrete objects *qua* data may only exist after humans get involved, but concrete objects *simpliciter* exist independently of whether they are taken as potential evidence or not.<sup>8</sup>

Adrian Currie (2021) discusses the data-time/phenomena-time distinction in the context of paleontology. He characterizes phenomena-time as “the lifetimes of the natural processes, entities and events that scientists seek to understand” which “draws our attention to information-loss due to natural historical processes,” as opposed to data-time which “highlights information-destroying processes within science itself” (Currie 2021, 105). For a fossil, this makes it sound as though phenomena-time encompasses everything up to the moment where investigators get involved, and data-time encompasses everything thereafter. However, as pointed out in section 2, it is not necessarily obvious that we should draw the line between phenomena- and data-time here. An alternative would be to say that the epistemic target of paleontology – the events of the deep past – occur during phenomena-time, and that data-time begins with the burial or fossilization of the traces of the past. One could also allow for a (substantial) temporal gap between phenomena-time and data-time. Or, as Currie (2024) points out, perhaps we need a context-sensitive answer to the question of where phenomena-time stops and data-time begins; this project may be aided by adding other time dimensions such as scientist-time and specimen-time, as suggested by Wylie (2024).

8. I focus herein only on recent philosophical accounts of data, because philosophy of data has been somewhat revamped in recent years. However, it is worth mentioning that Bogen and Woodward (1988), whose account of data/phenomena somewhat inspired Leonelli (2018)’s distinction between data-time and phenomena-time, also characterize the distinction between data and phenomena ambiguously. On the one hand, they say that data and phenomena have distinct epistemic roles; specifically, data serve as evidence for theories and theories (attempt to) explain phenomena. However, they also characterize the distinction in terms of observation – data are “uncontroversially observable” (Bogen and Woodward 1988, 314), whereas phenomena largely “cannot be perceived” (Bogen and Woodward 1988, 350). So, again, their account is ambiguous about whether data are best characterized by the role they play in inquiry or by their production by human activities such as measurement, and their account is silent concerning what we should think if and when these two features of data come apart.

In summary, it is currently unclear whether data should be characterized by their epistemic function or by their means of production by researchers. In other words, although becoming data is certainly a *relational* process, what does an object have to relate to in order to become data: a claim or question, or a researcher? Fossil data illustrate this ambiguity particularly well. Next, I elaborate on some remaining plausible views of data, and, accordingly, call for future refinement of the related concepts of data, data models, data journeys, metadata, and data-time.

## 4 Resolving the Ambiguity

Resolving the ambiguity described in section 3 will require more carefully defining five different technical terms: data, data journey, metadata, data-time, and phenomena-time.<sup>9</sup> For each term, we need to decide to define it in terms of the right kind of epistemic function (e.g., evidential role, representational role), or the right kind of human-world interaction (e.g., discovery, measurement), or both. Of course, we might also decide to define these terms differently in different scientific or philosophical contexts.

All of the various options are summarized in table 1.<sup>10</sup>

First, there are three options for defining the moment at which an object becomes a piece of *data*. We might say that an object becomes a piece of data if that object serves the right kind of epistemic function. Leonelli's definition of data as portable, potential evidence, for example, seems to follow this route, some of the complexities explored above notwithstanding. Alternatively, we might say that an object becomes a piece of data if it is generated by the right kind of human-world interaction, such as measurement or observation. Or, we could require both of these conditions – the right kind of epistemic function and the right kind of human-world interaction – before we say that an object counts as data (Bokulich and Parker [2021] seem to go this route).

The same three options could also apply to when we should start calling a process a *data journey*, as opposed to any other kind of process. Data journeys could occur to an object only after that object acquires a certain epistemic function, or only after that object is produced by the right kind of human-world interaction, or both. Importantly, and unlike in the definition of data journeys given by Leonelli (2020), which references data's "production site," data journeys need not apply to all and only objects that are data; we might decide to resolve the ambiguity about what counts as data differently from how we resolve the ambiguity about when data journeys begin. For example, data journeys may apply to an object before it becomes data (i.e., if there are more criteria for when something becomes data than there are for when an object can undergo data journeys).

Third, we could likewise define *metadata* as contextual information about objects that are taken to serve a particular epistemic function, or objects that are produced by certain human-world interactions, or both. Again, in principle, metadata could be defined using different criteria than are used to define data (or data journeys).

9. "Data model" might be added as a sixth term here. However, I've decided not to include it for the reason that philosophers of data so far disagree about the distinction between data and data models, in part because they disagree about the epistemic function of data (is it evidential or representational?). As I am attempting to *not* weigh in on the epistemic function of data in this paper, I've decided to also not weigh in on the data/data model distinction. Further work is needed to clear this up.

10. Technically, there is also the possibility of just rejecting existing philosophical accounts of data altogether and focusing on neither human-world interaction nor epistemic function. A large shortcoming of this option is that it rejects the enormous progress made in recent years in philosophy of data, so I won't entertain this option further.

	Epistemic function	Human-world interaction	Both
Data	An object becomes data as soon as it serves the right kind of epistemic function.	An object becomes data as soon as it is produced via the right kind of human-world interaction.	An object becomes data as soon as it is both produced by the right kind of human-world interaction <i>and</i> serves the right kind of epistemic function.
Data journey	An object can undergo data journeys after that object serves the right kind of epistemic function.	An object can undergo data journeys after it is produced via the right kind of human-world interaction.	An object can undergo data journeys after that object is both produced by the right kind of human-world interaction <i>and</i> serves the right kind of epistemic function.
Metadata	Metadata is information about processing affecting objects that serve the right kind of epistemic function.	Metadata is information about objects that were produced via the right kind of human-world interaction.	Metadata is information about objects that both were produced via the right kind of human-world interaction <i>and</i> serve the right kind of epistemic function.
Data-time	An object is in data-time once that object serves the right kind of epistemic function.	An object is in data-time once that object has been produced via the right kind of human-world interaction.	An object is in data-time once that object both has been produced via the right kind of human-world interaction <i>and</i> serves the right kind of epistemic function.
Phenomena-time	An object is in phenomena-time prior to when that object is used to serve the right kind of epistemic function.	An object is in phenomena-time prior to when that object undergoes the right kind of human-world interaction.	An object is in phenomena time prior to both when that object undergoes the right kind of human-world interaction <i>and</i> when it is used to serve the right kind of epistemic function.

Table 1: Three options each for how to define data, data journeys, metadata, data-time, and phenomena-time, in terms of whether these definitions should emphasize an object's epistemic function, its production via human-world interaction, or both. My preferred option for each term is highlighted.

Fourth, *data-time* could be defined in terms of the time at which an object acquires an epistemic function, the time at which a human-world interaction produces an object (as Leonelli [2018] currently defines it), or both; and, finally, *phenomena-time* could also be defined in terms of what objects with a certain epistemic function are used to do (as in Leonelli [2018]), what objects produced by certain human-world interactions are used to do, or both. The definitions of data- and phenomena-time need not be such that there is a clear or sudden transition from one to the other. For example, there could be a gap between phenomena-time and data-time (e.g., if there are more criteria for data-time beginning than for phenomena-time ending), or overlap between the two (e.g., if there are more criteria for phenomena-time ending than for data-time beginning).

With three options each for five different terms, we are left with fifteen different accounts (not even counting the various ways we might spell out the “right kind” of epistemic function or human-world interaction). Clearly more work needs to be done to evaluate all of the possibilities in depth; the primary contribution of this paper is to encourage this further work. However, I’ll close the paper by indicating which combination of views I find particularly plausible, in light of the fossil data case.

According to my preferred view: *data* are defined in terms of both epistemic-function and human-world interaction, *data journeys* are defined in terms of epistemic function, *metadata* are also defined in terms of epistemic function, *data-time* is defined in terms of both epistemic function and human-world interaction, and *phenomena-time* is defined in terms of epistemic function only. (These preferred options for each term are highlighted in table 1.) And here is what that would say about fossil data: A fossil specimen wouldn’t become a piece of *data* until later on, at least post-discovery (and maybe even later; this would be number “1” or “3” in figure 1). This is to prevent the counter-intuitive possibility that an object would count as data even if it was never used as such (and maybe was never even known by researchers to exist). However, the processes that affect fossil data but occur before humans get involved could reasonably be called *data journeys*, and information about those processes can be called *metadata*, consonant with contemporary paleontological practice and in recognition of the structural similarities in pre- and post-discovery processes detailed in section 2. The time prior to an organism’s death and burial would count as *phenomena-time* (this is the time period for which the later fossil data serve as evidence for paleontologists<sup>11</sup>), but only the time after a specimen counts as data would count as *data-time*. Defining phenomena-time and data-time this way leaves a (possibly very large!) gap between phenomena-time and data-time, a period yet to be labeled by philosophers of data or philosophers of paleontology.

## 5 Conclusion

This paper has used an examination of fossil data and the processes shaping them to illustrate an important and as-yet unresolved ambiguity in philosophy of data, concerning whether terms such as “data,” “metadata,” “data journey,” and “data-time” should be defined primarily by referring to the particular epistemic function an object plays in inquiry (e.g., its evidential role) or by how that object is produced or interacted with by researchers (e.g., by measurement). Fossil data are shaped by many processes that occur before humans ever encounter fossil specimens, and therefore put pressure on accounts that suggest that terms like metadata and data journeys can only apply after investigators are involved. Although I have offered what I take to be a plausible resolution of the ambiguity in section 4, I leave open the possibility that other, competing ways

11. For other scientists, like taphonomists, who specifically study fossilization processes, other time periods might count as phenomena-time. Currie (2024) argues for a similar, context-based definition of phenomena-time.

of resolving the ambiguity might be developed moving forward, and encourage other philosophers of data to explore these options by applying them to novel case studies.

## Acknowledgments

This paper benefited greatly from feedback by Laura Bianchi, Federica Bocchi, Alisa Bokulich, Julia Bursten, Matilde Carrera, Leticia Castillo Brache, Adrian Currie, Marina DiMarco, Doug Erwin, Caleb Hazelwood, Sabina Leonelli, Meghan Page, Emily Parke, Lydia Patton, Caitlin Wylie, Annette Zimmermann and two helpful, anonymous referees, as well as attendees of ISHPSSB in 2021, the “Beyond Incompleteness: New Perspectives on Fossil Data” SPSP cognate session at the PSA in 2022, the Philosophy of Science Reading Group at the Waipapa Taumata Rau University of Auckland in 2023, and the Data and AI Ethics Reading Group at the University of Wisconsin-Madison in 2023. This material is based upon work supported by the National Science Foundation under Grant No. (DGE-1840990). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The icons in Figure 1 include:

- extinction by Tippawan Sookruay from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/extinction-3385374/>.
- rock by Georgiana Ionescu from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/rock-2018151/>.
- soil profile by DracoVisions from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/soil-profile-5912519/>.
- tectonic by Nithinan Tatab from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/tectonic-2048518/>.
- science magnifying glass fossil by Maxicons from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/science-magnifying-glass-fossil-2186120/>
- Pickaxe by Eucalyp from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/pickaxe-4641828/>.
- shell fossil by Maxicons from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/shell-fossil-5380557/>.
- fossil bones by Template from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/fossil-bones-2873682/>.
- Velociraptor by Amethyst Studio from Noun Project; (CC BY 3.0); <https://thenounproject.com/icon/velociraptor-4131174/>.

## Literature cited

- Alroy, J., C. R. Marshall, R. K. Bambach, K. Bezusko, M. Foote, F. T. Fürsich, T. A. Hansen, et al. 2001. “Effects of Sampling Standardization on Estimates of Phanerozoic Marine Diversification.” *PNAS* 98 (11): 6261–66. <https://doi.org/10.1073/pnas.111144698>.
- Ankeny, Rachel, and Sabina Leonelli. 2021. *Model Organisms*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108593014>.
- Benton, Michael J. 2019. *The Dinosaurs Rediscovered: How a Scientific Revolution Is Rewriting History*. Thames/Hudson Limited.
- Bogen, James, and James Woodward. 1988. “Saving the Phenomena.” *Philosophical Review* 97 (3): 303–52. <https://doi.org/10.2307/2185445>.
- Bokulich, Alisa. 2021. “Using Models to Correct Data: Paleodiversity and the Fossil Record.” *Synthese* 198:5919–40. <https://doi.org/10.1007/s11229-018-1820-x>.

- Bokulich, Alisa, and Wendy Parker. 2021. "Data Models, Representation, and Adequacy-for-Purpose." *European Journal for Philosophy of Science* 11:31. <https://doi.org/10.1007/s13194-020-00345-2>.
- Boyd, Nora Mills. 2018. "Evidence Enriched." *Philosophy of Science* 85 (3): 403–21. <https://doi.org/10.1086/697747>.
- Cisneros, Juan Carlos, Nussaibah B. Raja, Aline M. Ghilardi, Emma M. Dunne, Felipe L. Pinheiro, Omar Rafael Regalado Fernández, Marcos A. F. Sales, et al. 2022. "Digging Deeper Into Colonial Palaeontological Practices in Modern Day Mexico and Brazil." *Royal Society Open Science* 9 (3): 210898. <https://doi.org/10.1098/rsos.210898>.
- Currie, Adrian. 2021. "Stepping Forwards by Looking Back: Underdetermination, Epistemic Scarcity and Legacy Data." *Perspectives on Science* 29 (1): 104–32. [https://doi.org/10.1162/posc\\_a\\_00362](https://doi.org/10.1162/posc_a_00362).
- Currie, Adrian. 2024. "Fossils, Modality & Central Subjects in Palaeobiological Reconstruction." *Philosophy, Theory, and Practice in Biology* 16 (2): 6. <https://doi.org/10.3998/ptpbio.5287>.
- Dean, Christopher D., Philip D. Mannion, and Richard J. Butler. 2016. "Preservational Bias Controls the Fossil Record of Pterosaurs." *Palaeontology* 59 (2): 225–47. <https://doi.org/10.1111/pala.12225>.
- Finkelman, Leonard. 2019. "Crossed Tracks: Mesolimulus, Archaeopteryx, and the Nature of Fossils." *Biology & Philosophy* 34 (2): 28. <https://doi.org/10.1007/s10539-019-9680-4>.
- Holland, Steven M. 2016. "The Non-Uniformity of Fossil Preservation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1699): 20150130. <https://doi.org/10.1098/rstb.2015.0130>.
- Jeffares, Ben. 2010. "Guessing the Future of the Past: Derek Turner, *Making Prehistory: Historical Science and the Realism Debate*." *Biology & Philosophy* 25 (1): 125–142. <https://doi.org/10.1007/s10539-009-9155-0>.
- Lautenschlager, Stephan. 2016. "Digital Reconstruction of Soft-Tissue Structures in Fossils." *The Paleontological Society Papers* 22:101–17. <https://doi.org/10.1017/scs.2017.10>.
- Leonelli, Sabina. 2015. "What Counts as Scientific Data? A Relational Framework." *Philosophy of Science* 82 (5): 810–21. <https://doi.org/10.1086/684083>.
- Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.
- Leonelli, Sabina. 2018. "The Time of Data: Timescales of Data Use in the Life Sciences." *Philosophy of Science* 85 (5): 741–54. <https://doi.org/10.1086/699699>.
- Leonelli, Sabina. 2019. "What Distinguishes Data From Models?" *European Journal for Philosophy of Science* 9 (2): 22. <https://doi.org/10.1007/s13194-018-0246-0>.
- Leonelli, Sabina. 2020. "Learning From Data Journeys." In *Data Journeys in the Sciences*, edited by Sabina Leonelli and Niccolò Tempini, 1–24. Springer. [https://doi.org/10.1007/978-3-030-37177-7\\_1](https://doi.org/10.1007/978-3-030-37177-7_1).
- Lescaze, Zoë. 2017. *Paleoart: Visions of the Prehistoric Past*. Taschen.
- Liboiron, Max. 2021. "Decolonizing Geoscience Requires More Than Equity and Inclusion." *Nature Geoscience* 14 (12): 876–77. <https://doi.org/10.1038/s41561-021-00861-7>.



- Lloyd, Graeme T., Paul N. Pearson, Jeremy R. Young, and Andrew B. Smith. 2012. "Sampling Bias and the Fossil Record of Planktonic Foraminifera on Land and in the Deep Sea." *Paleobiology* 38 (4): 569–84.
- Lloyd, Graeme T., Jeremy R. Young, and Andrew B. Smith. 2012. "Taxonomic Structure of the Fossil Record Is Shaped by Sampling Bias." *Systematic Biology* 61 (1): 80. <https://doi.org/10.1093/sysbio/syr076>.
- Mayor, Adrienne. 2013. *Fossil Legends of the First Americans*. Princeton, NJ: Princeton University Press.
- Mitchell, W. J. T. 1998. *The Last Dinosaur Book: The Life and Times of a Cultural Icon*. Chicago: University of Chicago Press.
- Monarrez, Pedro M., Joshua B. Zimmit, Annaka M. Clement, William Gearty, John J. Jacisin, Kelsey M. Jenkins, Kristopher M. Kusnerik, et al. 2021. "Our Past Creates Our Present: A Brief Overview of Racism and Colonialism in Western Paleontology." *Paleobiology*, 1–13. <https://doi.org/10.1017/pab.2021.28>.
- Noble, Brian. 2016. *Articulating Dinosaurs: A Political Anthropology*. Toronto: University of Toronto Press.
- Nyhart, L. K. 2004. "Science, Art, and Authenticity in Natural History Displays." In *Models: The Third Dimension of Science*, edited by S. de Chadarevian and N. Hopwood, 307–36. Stanford, CA: Stanford University Press. <https://doi.org/10.1515/9781503618992-014>.
- Peters, Shanan E. 2005. "Geologic Constraints on the Macroevolutionary History of Marine Animals." *PNAS* 102 (35): 12326–31. <https://doi.org/10.1073/pnas.0502616102>.
- Peters, Shanan E., and Michael Foote. 2001. "Biodiversity in the Phanerozoic: A Reinterpretation." *Paleobiology* 27 (4): 583–601. [https://doi.org/10.1666/0094-8373\(2001\)027<0583:BITPAR>2.0.CO;2](https://doi.org/10.1666/0094-8373(2001)027<0583:BITPAR>2.0.CO;2).
- Raja, Nussaibah B., Emma M. Dunne, Aviwe Matiwane, Tasnuva Ming Khan, Paulina S. Nätscher, Aline M. Ghilardi, and Devapriya Chattopadhyay. 2022. "Colonial History and Global Economics Distort Our Understanding of Deep-Time Biodiversity." *Nature Ecology & Evolution* 6 (2): 145–54. <https://doi.org/10.1038/s41559-021-01608-8>.
- Raup, David M. 1972. "Taxonomic Diversity During the Phanerozoic." *Science* 177 (4054): 1065–71. <https://doi.org/10.1126/science.177.4054.1065>.
- Raup, David M. 1976. "Species Diversity in the Phanerozoic: An Interpretation." *Paleobiology* 2 (4): 289–97.
- Rieppel, Lukas. 2019. *Assembling the Dinosaur: Fossil Hunters, Tycoons, and the Making of a Spectacle*. Cambridge, MA: Harvard University Press.
- Shipman, Pat. 1981. *Life History of a Fossil: An Introduction to Taphonomy and Paleocology*. Cambridge, MA: Harvard University Press.
- Smith, Andrew B. 2001. "Large-Scale Heterogeneity of the Fossil Record: Implications for Phanerozoic Biodiversity Studies." *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 356 (1407): 351–67. <https://doi.org/10.1098/rstb.2000.0768>.
- Smith, Andrew B., and Alistair J. McGowan. 2005. "Cyclicality in the Fossil Record Mirrors Rock Outcrop Area." *Biology Letters* 1 (4): 443–45. <https://doi.org/10.1098/rsbl.2005.0345>.
- Turner, Derek D. 2007. *Making Prehistory: Historical Science and the Scientific Realism Debate*. Cambridge: Cambridge University Press.

- Turner, Derek D. 2016. "A Second Look at the Colors of the Dinosaurs." *Studies in History and Philosophy of Science* 55:60–68. <https://doi.org/10.1016/j.shpsa.2015.08.012>.
- Turner, Derek D. 2019. "Speculation in the Historical Sciences." *Philosophy, Theory, and Practice in Biology* 11:11. <https://doi.org/10.3998/ptpbio.16039257.0011.011>.
- Vinther, Jakob, Derek E.G. Briggs, Richard O. Prum, and Vinodkumar Saranathan. 2008. "The Colour of Fossil Feathers." *Biology Letters* 4 (5): 522–25. <https://doi.org/10.1098/rsbl.2008.0302>.
- Wylie, Caitlin Donahue. 2016. "Overcoming Underdetermination." <http://www.extinctblog.org/extinct/2016/4/11/overcoming-underdetermination>.
- Wylie, Caitlin Donahue. 2019. "Overcoming the Underdetermination of Specimens." *Biology & Philosophy* 34 (2): 24. <https://doi.org/10.1007/s10539-019-9674-2>.
- Wylie, Caitlin Donahue. 2021. *Preparing Dinosaurs: The Work Behind the Scenes*. Cambridge, MA: MIT Press.
- Wylie, Caitlin Donahue. 2024. "Timing Science: The Temporal Role of Scientists in the Construction of Data." *Philosophy, Theory, and Practice in Biology* 16 (2): 8. <https://doi.org/10.3998/ptpbio.5646>.

---

© 2024 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits anyone to download, copy, distribute, display, or adapt the text without asking for permission, provided that the creator(s) are given full credit.

ISSN 2475-3025