Snails, time, data: On the politics of mass-digitization and the possibility of data drift

Roos Hopman (iD)

Abstract

In natural history museums around the globe, museum staff are working on making collections of specimens (such as insects on pins, snail shells, dried plants, mounted birds) digitally available through online databases en masse. The mass-digitization of specimens is urgent especially now, so museum actors and biodiversity scientists stress, as natural history collections could provide the biodiversity sciences with crucial data for research into the effects of climate-change induced biodiversity loss. This paper investigates the processes and practices through which these biodiversity data are produced. It argues that the work of data production requires the negotiation of- and grappling with different temporalities, and that attending to these temporalities contributes to a further understanding of the politics of mass-digitization. Building on ethnographic fieldwork in the Museum für Naturkunde Berlin's collection of snail shells, the paper brings into view the work required to make snail specimens and their metadata digital. Following steps in a mollusk digitization workflow—meant to streamline and scale up digitization processes—it attends to the misalignment of rhythms; the importance of informal cleaning labor; the negotiation of social, biological, and colonial temporalities in practices of label transcription; and the anticipatory and expansive logics of mass-digitization. Drawing on snails' ability to "drift," it raises data drift as a way of engaging with natural history collections and their data.

Keywords

Data production, mass-digitization, temporal politics, natural history, biodiversity, snails

Kingdoms of dark data

In this paper, I think through the production of digital data while considering an unlikely figure: the snail. This does not mean presenting a story about the fast and the slow. Rather, it means learning that iterations of pace, despite speed being evoked as a central rationale for digitization in memory institutions and beyond, might not be the most helpful analytical perspective when contemplating data production (and snails). But for now, let us go back to October 2021, when a machine affectionately called "the Beast" ("It's big, and it's loud"),[1] arrived at Berlin's natural history museum, the Museum für Naturkunde (hereafter: MfN). As part of a "unique pilot project," this Beast, the Entomology Conveyor, was to enable the industrial-scale digitization of the museum's insect collection. The system, run by human scan operators, was to carry pinned insects and the labels attached to them over a conveyor belt, photographing them in high resolution along the way. The photographs of the insects would then be saved on a server together with their metadata and eventually made "digitally

accessible"[2] to science and society through an online infrastructure.

In press releases published ahead of its arrival, the museum's leadership stressed that this novel system would significantly speed up the process of rendering its collections digital: "Now, cost-effective, fast, and accurate automated digitization is within reach: By using the new and unique Entomology Conveyor, the Museum für Naturkunde Berlin will be able to massively speed up the digitization process."[3] The rhetoric of "speeding-up" presented here is central to arguments made in favor of digitizing natural history collections and making biodiversity data

European Ethnology, Humboldt Universität zu Berlin/Museum für Naturkunde Berlin, Berlin, Germany

Corresponding author:
Roos Hopman, European Ethnology, Humboldt Universität zu Berlin/ Museum für Naturkunde Berlin, Anton-Wilhelm-Amo-Straße 40-41, Berlin, Germany.
Email: hopmanro@hu-berlin.de

more accessible. Researchers in natural history and (evolutionary) biology have emphasized the need to digitize on an unprecedented scale, to massively "scale up" and "accelerate" the process of making collections digital (see, e.g. Hedrick et al., 2020; Heerlien et al., 2015). This urgency is often linked, I observed, to a need to produce data for biodiversity and environmental research. Indeed, this literature regards specimens in natural history collections as untapped sources of data—they are therefore also referred to as "dark data" (see, e.g. Marshall et al., 2018). As the Director of the Smithsonian Museum of Natural History recently remarked about natural history collections in the New York Times, "It occurred to us that we each hold these amazing assets … We realized that we were presiding over these kingdoms of dark data."[4] Once there is open access to historical specimen data, the argument goes, scientists will be able to make comparisons of species distribution over time, potentially helping to slow down or indeed halt climate-induced biodiversity losses. At the same time, the emphasis on mass-digitization draws attention to the role played by temporality in the production of digital data more generally. Not incidentally, when the term "big data" was first coined in the early 2000s, velocity was, alongside volume and variety, one of its key attributes (Kitchin and McArdle, 2016).

In this paper, I consider the project of mass-digitizing natural history collections through the lens of data production. I deliberately use the term "data production" here: I am not so much interested in practices surrounding the *processing* of data, that is, preparing them for analysis (see, e.g. Leonelli, 2019; Plantin, 2021), but more in the translation of physical objects into digital forms. In part, this is a byproduct of the institutional setting in which I conducted my research: for this natural history museum, the effort of transforming physical specimens into digital formats seemed to come before concerns about how the resulting data might subsequently be stored, managed, used, and maintained. Perhaps comparable to "big biology" initiatives (Davies et al., 2013) or border-control practices (Leese and Pollozek, 2023), biodiversity science is a field where the sheer *availability* of ever-more data is expected to contribute to the betterment of research and governance (in this case, through conservation and ecosystem management measures). However, the socio-material conditions of data production matter to how data is then processed and made available, and as such to how biodiversity crises are approached and managed. As the above already suggests, these are not purely technical processes (Thylstrup, 2018). I will show how the work of data production requires those digitizing to negotiate and grapple with different temporalities, arguing that attending to these differences can contribute to a further understanding of the politics of mass-digitization (Thylstrup, 2018). Paying attention to temporality offers insights into the misalignment of rhythms; the importance of informal cleaning labor; the negotiation of

social, biological, and colonial temporalities in practices of label transcription; and the anticipatory and expansive logics of mass-digitization.

My analysis builds on two years of working at the MfN as a guest researcher between 2021 and 2023, a period in which I collaborated on a project studying museums as spaces of social cohesion. The main focus of my project was the then recently-launched mass-digitization endeavor, in particular, the discursive and material practices involved in the project. What were the envisioned uses and promises of these data, and why did there appear to be such an urgency to produce them en masse? To find out, I attended museum meetings on the various digitization projects and attended conferences on biodiversity digitization. I was shown around by curators and conducted interviews with digitization managers, scan operators, and engineers in order to understand the rationales, promises, and complications of (generating) natural history collection data. As an anthropologist of scientific practice, I wanted to contextualize my findings through the workings of digitization "on the ground." I thus conducted three months of focused ethnographic fieldwork in the malacology (mollusk) collection. While my attention was initially drawn to the Entomology Conveyor, one digitization manager suggested I do my research with the mollusks instead. He expressed his concern that the Entomology Conveyor project would "overshadow" other digitization projects at the museum and told me that they were "always in need of help and hands"[5] in malacology. I worked in the collection two to three days a week, transcribing specimen labels and scanning shells along with my collection-based colleagues, who were simultaneously my interlocutors.

Encompassing an estimated 5 to 8 million individual snails (or rather, their shells), the malacology collection is the museum's largest after the insect collection,[6] and is thus considered a pivotal testing ground for digitization processes—and not only for this institution. With their partly automated setup for digitizing snail shells, my colleagues in malacology were hoping to set a precedent for other institutions aiming to digitize their collections (including, e.g. ancient papyrus documents). Mollusks, the taxonomic group to which snails belong, are furthermore deemed important indicators of the species diversity and health of local ecosystems; they are considered "sentinel species" (Van Dooren, 2022: 201). Their habitats are often small, and because of their sensitivity to moisture and temperature, they do not adapt well to changes to their environment. Their "distribution in time and space" therefore provides "a foundation for [the] environmental monitoring of all human-impacted habitats" (Sierwald et al., 2018: 177).

To be sure, the extinction of species and the destruction of ecosystems are posing crises that are real and demand urgent responses. However, the proposed response of making biodiversity data available begs critical engagement, as presenting data as a solution and digitization as a

linear, easily scalable project sidelines questions about how the historicity and socio-material conditions (Bates et al., 2016) of digitization shape the resulting data (Kaiser et al., 2023). In the following, I would like to draw attention to practices of digitizing snail shells, in particular the human labor of cleaning and transcription it requires. I will be guided by the steps specified in the malacology collection's digitization workflow, introduced to make the work smoother and more efficient, showing that digitization is a process not easily streamlined or scaled up. In the conclusion, inspired by snails' ability to "drift" (Van Dooren, 2022: 61), I will raise the possibility of what I call "data drift" as a way of relating to natural history collections and their data.

Digitization and data practices

In her work on digitization in the cultural heritage sector, Nanna Bonde Thylstrup (2018) offers an initial understanding of mass-digitization and its politics. While not specifically addressing natural history and its collections, her work proves instructive when considering snails and other to-be-digitized life-forms. Thylstrup reminds us that mass-digitization is never "a neutral technical process" (2018: 5), suggesting to approach "mass digitization projects as … emerging sociopolitical and sociotechnical phenomena that introduce new forms of cultural memory politics" (2018: 4). She argues that these new memory politics are in part a result of the enrolment of a new host of actors (and with that, new interests) in otherwise public institutions. It is precisely the "conflicting motives and dynamics" emerging from these (to a large extent commercial) actors and interests that Thylstrup identifies as key to analyzing the politics of mass-digitization. She locates these tensions above all on the infrastructural level, thus calling for an "infrapolitics" of mass-digitization.

Thylstrup's impressive work recognizes the importance of "everyday practices" when considering infrapolitics, but these practices remain somewhat abstract in her descriptions; the main actors in her analyses and stories are public institutions such as libraries and museums, commercial companies, and platforms such as Google, as well as smaller-scale digitization initiatives. I add further nuance and concreteness to Thylstrup's work by offering a detailed ethnographic account of mass-digitization in practice, paying special attention to the dimensions of time and temporality. My own take on politics, then, is inspired by work in science and technology studies and actor-network theory (ANT) in particular, which approaches the worlds around us as complex networks of heterogeneous relations. ANT is not so much concerned with the explicit *motives* and *dynamics* of human actors but rather focuses on *how things are done*: their practices (on digitization, see, e.g. Roth and Bowen, 1999; Shavit and Griesemer, 2011). Following this approach, politics emerge not necessarily in conflicting interests, but in the differential effects produced through the ways that things, people, processes, temporalities, and materials come to relate in practice.

As suggested above, I characterize the practices considered here as practices of *data production*. Philosopher Sabina Leonelli's work has provided detailed insights into a wide variety of data practices, or "activities involved in handling data" (Leonelli, 2020: v). In these descriptions, *data production* and *data processing* are often taken to refer to the same set of practices. I aim to add to Leonelli's analyses by distinguishing data production as a separate set of practices. Whereas *data processing* can be characterized as the cleaning, organizing, and structuring of data (see, e.g. Boumans and Leonelli, 2020; Plantin, 2021) in order to prepare them for analysis, *data production* in this case revolves not so much around the *handling* of data, but rather the socio-material conditions in which data *come into being*. We find examples of this in Leonelli's work on phenomics (2019), where she details, for example, "plants playing tricks" (p. 22) by dropping leaves during digitization processes, the demands of soil and lighting, and the quirks of digitization systems. Where other analyses of mass-digitization have paid more attention to the processing, movement, and use of data—for example, the value of mass-digitized content for different users—or the novel demands placed on (cultural) heritage institutions to acquire computational expertise and "data skills" (Ames and Lewis, 2020; Bonacchi, 2021), I offer a detailed account of data production in particular.

Finally, where we are matters. As outlined in the introduction, data from natural history specimens is being produced against the backdrop of biodiversity loss. Mass-digitization is seen to serve the development of "global databases" of biodiversity data (Nadim, 2021), or what might be referred to as "biodiversity panopticons" (Bowker, 2000: 645). These data are generally taken to inform biodiversity research, "to enhance our understanding of biodiverse systems" (Nelson and Ellis, 2018: 1). As ecologists Anne Magurran and Brian McGill write, "Climate change and biodiversity loss have increased the urgency for more and better data" (2011). By studying these practices, I therefore additionally contribute to scholarship on the politics and "datafication" of nature (Nadim, 2021).

Snails have shaped the very digitization processes and technologies that we encounter in the museum collection: both the development and eventual form taken by the scanning machine—which was adapted specifically to accommodate the diversity of snail shells—and the way the work is organized, as those digitizing were deliberately physically removed from the collections to prevent them from becoming too enthralled in the shells (i.e. to avoid the risk of the snails eating away at their work time and efficiency). Indeed, as noted by Marisa Anne Bass and colleagues in their aptly titled edited volume *Conchophilia*,

"Shells have long fascinated through their variety, their beauty, their shapes, colors, and luminescence" (2021: 3). In this sense, these snails (and nature more generally) are not simply passive objects in digitization processes. Rather, as suggested by Sarah Besky and Alex Blanchette (2019: 2–3), they invite us to consider that "the nonhuman world" has the "capacity to work upon us, against us, and perhaps with us."

How can snails, then, help us to think about data production and temporality? Snails are often reduced to caricatures of slowness, which, observing how they move, is not wholly surprising. Yet, snails have an astonishing ability to travel great distances, to which their presence on islands around the world attests. Snails can close off their apertures with a mucus membrane, allowing them to float in salty water for weeks on end, and scientists have theorized that they have reached remote islands by crawling onto the backs of birds. Field philosopher Thom van Dooren therefore sees them as capable of "a particular form of deep time travel: *drifting*." As he goes on to summarize, "snails amaze with their capacity to move so far, to spread so widely, while doing so little … the highly successful passivity of snails might be seen as a remarkable evolutionary achievement" (2022: 61). It is snails' capability to forego control and direction, to *drift,* that I would like to take seriously in this paper.

From a management perspective, drifting and drifters tend to be undesirable. Drifting implies "a loss of direction," "a passive and listless movement, an aimless wandering and driving astray; a loosing [sic] of path and meaningful course. The drifter, tramp, ranger, vagabond, is seen as unpredictable and vague" (Pétursdóttir, 2020: 98). Rather than constituting failure or deviance, something that should be avoided or sanctioned, I argue that drifting might instead offer a means of engaging with different temporalities and, thus, the politics of mass-digitization.

Digital snails: Data production in a natural history museum

As a centralized place for the collecting and ordering of nature and natural things, the natural history museum has long been a site for "rendering worlds and their inhabitants into (digital) data" (Nadim, 2021: 67). Since the mid-twentieth century and with the advent of new database technologies and data storage devices (Bowker, 2005: 132), natural history museums' "memory practices" have become increasingly focused on digital formats. In 1971, for example, zoologist Peter J.P. Whitehead, who himself specialized in the study of fish, argued for the necessity of a "quick retrieval system" for "reviewing the holdings of any species in any museum" (1971: 219): "Only a computer can keep systematic information up-to-date and easily retrievable" (211). Indeed, the digitization of collections

has been a recurrent concern in museums since the 1970s (Parry, 2007). It is in light of these developments, and the growing urgency of biodiversity loss, that we should consider current efforts to digitize specimens.

Conducting fieldwork in the collections of the MfN, I found that the project of mass-digitization is to a large extent a logistical one (see also Nadim et al., 2024) revolving around establishing *what* is in the collections by giving specimens digital identifiers such as inventory numbers and QR codes. During an interview, the head of digitization explained that digitization entails moving information that is currently only available *physically*, in the shape of physical specimens and the text written on their paper labels, into a digital format. He explained that it is intended to make collections navigable and specimens "searchable and findable." The promise is that, with the QR code, it will become possible to additionally connect all the data related to a specimen. QR codes were therefore also described to me as an "anchor" for the data: by holding metadata, images, and other information in place, we might say the QR code prevents data from drifting. Of course, both digitization technologies and the collections themselves are constantly developing, changing, and moving: specimens continuously arrive at the museum, buildings undergo renovation, curators retire, technologies break down or are replaced, standards are updated, data gets lost. Thus, as the museum's information manager remarked when I interviewed her early on in my project, "digitization is a never-ending process."

Each collection at the MfN had its own technologies, databases, and guidelines for digitizing specimens: the differing materialities and curatorial logics of each collection (and sometimes of individual specimens) demanded an adaptive, idiosyncratic approach. In the malacology collection, a "dry" collection of snail shells, digitization work was divided into two parts: inventorization on the one hand and the scanning of specimens and their labels on the other. Inventorization was "the first, and most important step in the digitization process," as a scan operator in the collection emphasized. Inventorization encompassed checking specimens' inventory numbers and transcribing the metadata (inventory number, taxonomic information, collection location) on accompanying labels into a customized data tool (Figure 1). It also included adding a QR code in paper form: a tiny piece of paper, cut out by hand by the person doing the inventorizing. It was not the individual snails but "lots," small boxes containing a number of shells, that received a digital identifier in the shape of a QR code. In the digitization process, each lot of shells (some containing over 50 shells, some only two or three) was considered a unit, making these little boxes key devices in this context (Nadim, 2020).

After inventorization, the shells and their labels were photographed using a (partly) automated scanner. Because the objective was to digitize the whole collection by
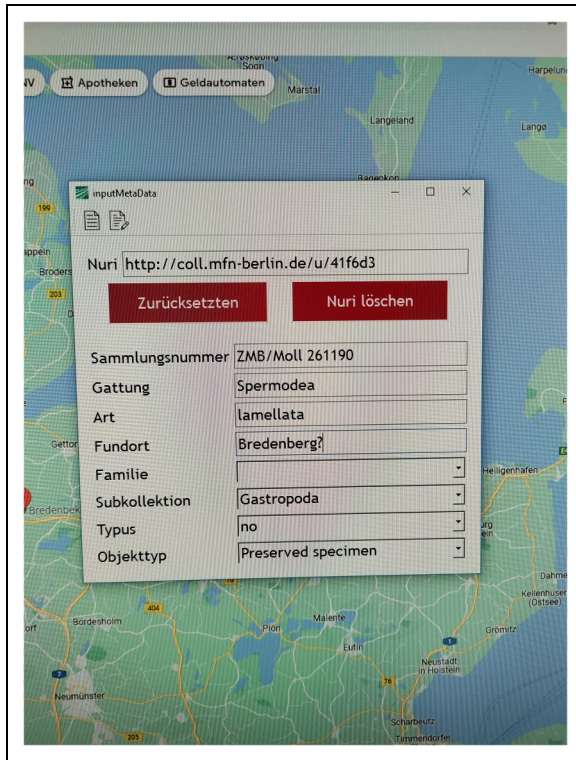
Figure 1. Metadata input interface.

2028, it was not feasible to photograph every individual shell. "That would take forever," the collection's digitization manager told me. Rather, if a lot contained one or two shells, they were photographed together in one shot, and if a lot contained more than three shells, the inventorizer selected three shells representative of that lot. The scanner used for this process was developed specifically for the digitization of snail shells and named "DORA" after the late, Berlin-based malacologist Dora Godan (1909–2006). The camera settings were specifically tinkered with to account for the wide variety of snail shells: their differing levels of shininess, their various colorings and shapes. A "scan operator" would place the shells onto a tray scattered with fine, dark sand and push it into the scanner. Four cameras then automatically took pictures of the tray from above, capturing the shells through apertural (the opening of the shell), dorsal (the back of the shell), and lateral (the side of the shell) shots (Callomon, 2019). Through this photographic process, the shells were standardized and reproduced as natural history specimens and objects for scientific inquiry (Daston and Galison, 2007: 134–135; Vennen, 2022), although considering the differing quality of these images, the question remains as to how they might actually be used to inform scientific research.

With the initial goal of digitizing this collection of seven million shells by 2028 in mind, one of the collection's scan o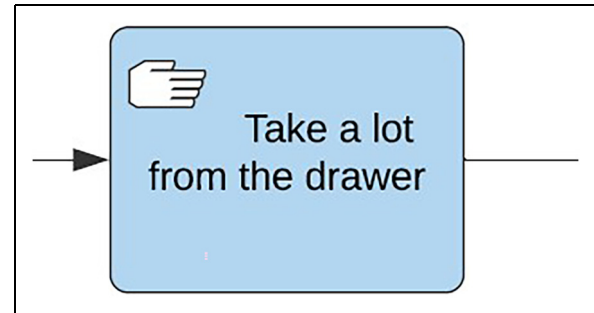perators had been tasked with developing a workflow to streamline the inventorization process. In the following section, I will make use of that workflow to turn my attention to digitization work. As the term suggests, a "workflow" is supposed to make the performance of work as smooth and efficient as possible. Dividing the digitization process into concrete steps to be completed by the inventorizer, it offers a linear sequence of tasks. However, with my fieldnotes, I will show that the digitization of specimens is a process not easily streamlined or ironed out in a linear fashion: it does not entail a simple translation from a physical to a digital format. I will draw particular attention to the act of cleaning and the transcription of specimen labels, as well as the anticipatory and expansive logics of digitization. As digitization processes were constantly changing, it is important to note that what I describe here exemplifies the processes as they played out in March, April, and May of 2022 (Figure 2).

Figure 2. Taken from "Digitation of the Mollusc collection. Input Metadata tool." March 2022.

Cleaning lots: Making sedimentary legacies legible

Taking one lot out of one of the hundreds of drawers is not the first but the third step described in the mollusk digitization workflow. By this stage, the inventorizer had already walked over to the wooden cabinets and pulled out a drawer filled with shells (step one) and then created a unique folder on their computer designating the cabinet and drawer (step two). These folders were where the files for each lot of shells were to be stored. The digitizer then had to reach into the drawer and select one of the lots containing shells, lined up inside the drawer like little bricks. Some of the drawers were filled to the brim with lots, whereas the inventorizer might only find a few in others. In some of the fuller drawers, the lots were scattered about haphazardly with no apparent order to them. To make sure none of the lots were forgotten, one of the scan operators explained, "you need to find an organized way to work through the drawer. It's like Tetris, you put everything where it fits." In addition, the original lot, made of cardboard, might have lost some of its shape and firmness in the period between collection and

digitization. To ensure the shells remained well-contained, digitizers in some instances would therefore put the shells into a new (plastic) box altogether. This process was also referred to as "rehousing" and shows how material transformations are crucial to processes of data production (Bates et al., 2016; Gitelman, 2013: 6; Wilson, 2011).

Since this collection of snails was not one where the specimens were looked at particularly frequently, most of the lots had accumulated a layer of dust and grime. In fact, my colleagues speculated that nobody had probably looked at these shells since they were first collected. This meant that most lots had been gathering dust for upwards of 100 years. Since the inventorizer was the first person to open the lots up and take a look at the shells inside, my colleagues and I would use a ball of cotton to wipe away any debris that had accumulated on the lot's glass lid. Making the lot look "clean" was partly about aesthetics, a pleasing task that added to the inventorizers' "job satisfaction" (Mol, 2020: 395). Moreover, it enabled the inventorizer to see inside the lot and determine whether they had already given it a QR code and, hence, to avoid digitizing the same lot twice. Dirt was not only a problem in the malacology collection. During a digitization tour through the mammal collection, the curator pointed out the importance of cleaning prior to digitization, saying, "We need to get rid of 150 years of dust to read what's on the skulls and the labels."

Besides dust, "dirt" took many other forms in this collection. I would, for example, encounter little pieces of cork, a material that had been used to stopper the glass tubes in which some of the smaller shells had been collected. Another frequent find were the exoskeletons of insects that had fed on any tissues left behind in the shell and that had since passed away. The amount of dead material found in the lots depended on the drawer: some shells seemed to be cleaner than others, meaning they had contained less residual tissue and had therefore attracted fewer insects. A colleague described the lots filled with these skeletons to me as "little graveyards." To a significant extent, cleaning was therefore about removing sediments and remnants that had gathered on- and inside these lots, and is thus a practice that bears witness to the passing of time. For the collection to be made digital, those sediments needed to be removed: both to render the texts on labels, containing taxonomic and collection information, readable, and to make the snails visible. To borrow from Shana Lee Hirsch, David Ribes, and Sarah Inman, cleaning here can consequently be seen as a manner of reckoning with the collection's "sedimentary legacies" (2022). Drawing on Star and Ruhleder (1996), Hirsch, Ribes, and Inman write that "historical accumulation" is characteristic of long-term infrastructures, which, they argue, "'sediment' over time": "an infrastructure is built 'on top of' its own past" (2022: 564). Through the handling and cleaning of each lot of snails, parts of the legacies of these collections were made visible and prepared for reproduction.

Removing sedimented materials from the lots or, in some cases, putting the shells in a new plastic container, was a time-intensive process, in particular because the shells were fragile and sometimes very small—some shells were not much bigger than a grain of sand. Workers did their best to avoid damaging the fragile shells in any way while working with them, for example, by handling them with tweezers. Sometimes they used softer, more flexible tweezers, which were especially suitable for picking up and transferring fragile shells, and sometimes harder ones, mainly used to take the firmer paper labels out of the lots. Despite all the care taken not to damage them, it was unavoidable that shells would sometimes "jump" out of the lot and fall onto the floor, forcing me to dive under my desk. The fear of damaging specimens affected how work was performed in the collections, but on a higher level it was also used as an important argument in favor of mass-digitization. Digitization holds the promise that fragile physical specimens will require less handling in the future, allowing them to be safely stored away. In this imaginary, like in Tetris, once a row of specimens has been digitized, they can "disappear."

Digitization, I therefore learned, involves a lot of cleaning—both in the sense of removing dirt and, with the "rehousing" of shells and the Tetris-like work of (re)organizing drawers, a more spatial practice of putting everything in its proper place (Mol, 2020). This extended to the scanning process: when taking pictures with DORA, I soon found that many of the plastic tubes containing shells were covered in a chalky white substance. This made it impossible to see (and hence photograph) the snails. Cleaning these tubes before photographing them thus became essential to capturing the shells digitally. While using DORA one time, I complained to a colleague about this. She acknowledged that dirt on the tubes was indeed a problem: "We have asked for new, clean materials many times, but haven't received any … Sometimes I just try to clean the tubes a bit myself. Actually, we are doing a bit of that work, cleaning and replacing the tubes. But it really should not be our job. That is collection caretaking." Part of the reason for this, my colleagues explained, was that requesting new materials was a slow and cumbersome process that could take months, with documents to be filled out, signatures to be gathered, and orders to be placed. "Just doing it yourself," then, became a more viable option for dealing with such issues. This demonstrates the different rhythms of digitization, in particular how the rhythms of scanning work were misaligned with the "organizational rhythm" or "institutional time" (Jackson et al., 2011) of the museum.

The cleaning work, while forming a substantial part of the inventorizers' and scan operators' tasks, was not recorded in the malacology workflow and therefore
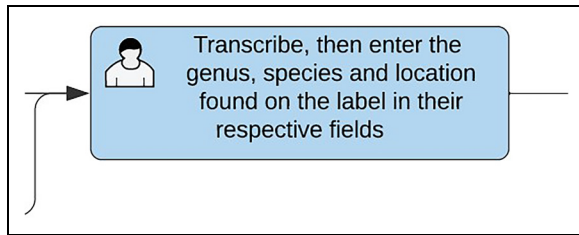
Figure 3. Taken from "Digitation of the Mollusc collection. Input Metadata tool." March 2022.

remains "unspecified and unrepresented" (Star and Strauss, 1999) in descriptions of digitization processes. Engaging with cleaning work, however, lays bare the different rhythms at work in digitization projects—rhythms that are not necessarily in sync. This dissonance signals a disconnect between upper management and digitization workers in the collections, causing frictions and delays in everyday work processes. This section has also shown that cleaning is a way of dealing with sedimented time, a practice where layers of dust attest to a history of disuse. The removal of this dust and other debris from the lots, the snails, and their labels revealed text and information written by collectors on a particular specimen or batch of specimens. Cleaning is thus literally a way of making the specimens' histories, inscribed on the labels, legible, and since it makes the shells visible to a scanner, it is imperative to capturing them digitally in the first place. Considering the presence of this dust and the apparent disuse of the shells, this again begs the question of how these snails, once digitized, might contribute to biodiversity science. Finally, cleaning specimens, or the "making of physical improvements,"[7] contributes to the continued preservation of specimens and, as such, to the role of the natural history museum as an institution of conservation (Figure 3).

Label travels: The imaginative work of transcription

Let us go a step further in the workflow. The inventorizer has added a QR code to the lot and has identified whether this lot already has an inventory number. It is now time to move on to the *label*, a piece of paper that is "like a strange, very brief, postcard" (Van Dooren, 2022: 129) from a collection site. The digitizer has to read and transcribe the information deemed most relevant (in this case: the genus, species, and collection location) from the label into the "metadata input tool" (Figure 1), a customized interface for recording label data.

While inventorizing, I quickly found that recording the *location* where specimens had been collected was the most time-consuming task. The location had often been scribbled down in handwriting incomprehensible to me (a non-expert future reader). In addition, since many of the shells had been collected in the early twentieth century,

locations were often written in Sütterlin, a script commissioned by Prussia in 1915. Sütterlin became the empire's official script, the "deutsche Volksschrift" [script of the German people] (Schopp, 2016: 286), and was taught in schools Germany-wide until 1941, when it was banned by the Nazi regime. Some Sütterlin letters look entirely different to those used in contemporary German, and identifying them takes time and skill. When one of the scan operators found me groaning over an unreadable label one day, she advised me to keep a tab with the Wikipedia page for Sütterlin open while inventorizing, which allowed me to compare the scripts directly. As she recounted, "At first, I constantly had to compare the writing on the labels to this image. But after a while you learn to recognize the letters." As such, I often found myself lifting a label to my computer screen between tweezers, squinting at the letters as I tried to figure out what location the collector might have meant. Take this excerpt from my fieldnotes:

> Deciphering the handwriting on labels is often challenging. It forces me to be creative: I open google and type in potential town names until I get a match that seems probable. And even when the handwriting is clear, the location may refer to a place that is now spelled in a different way. A label that had "Bredenberg" written on it is an example of this. There is no such town to be found on google maps, only a street. There is however a municipality called "Breddenberg" and another one called "Bredenbek." It might be either of these two places. The scan operator suggests I stick with the label, and add a question mark to it. (see Figure 1)

Due to the age of the labels, some of the locations noted on them simply no longer existed. Some had changed names, had been absorbed into neighboring towns and cities, or seemed to have disappeared into thin air altogether. Every label therefore presented a puzzle, requiring the inventorizer to both decipher the handwriting and search the internet and collection catalogues for clues as to the location the label's author might have been referring to. The transcription of labels thus requires historical curiosity and a willingness to investigate on behalf of the digitizer. At the same time, the location search evinces the different temporalities at play in collections: the biological temporalities of taxonomy and species-naming practices, the political temporalities demonstrated by shifting geographies and scripts, and the technological temporalities of collection catalogues and computers.

While the labels often contained information about a geographical location, they would sometimes also describe landmarks (a mill or bridge), the landscape (rivers, meadows), geological details (limestone, sandstone), metrical directions (e.g. 500 m from somewhere or at an elevation of 1000 m), or directions (such as east/west/south of something). These descriptions could get very detailed.
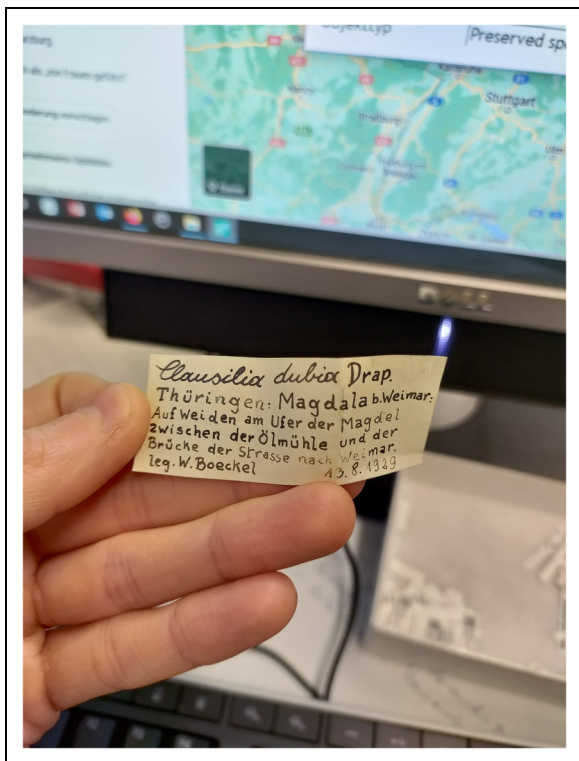
Figure 4. Descriptions.

Figure 5. "Gattung?" [Genus?].

One (see Figure 4), for example, reported that it had been collected "on pastures on the banks of the Magdel [river] between the oil mill and the bridge on the road to Weimar." What actually remained of this information after digitization was limited: it pertained to geographical locations in terms of urban settlements that had been recorded, not distances or landscape details. Inventorizers would summarize this information in the location field of the data tool as, for instance, "Magdel bei Weimar." In light of the mass-digitization goal, or processing the set number of specimens within the timeframe of the mass-digitization funding, such details became redundant. "The inventorization we perform here in malacology is minimal data. Minimal data is faster," the scan operator explained, "and it offers a quick overview for online users. They can quickly see what is there and decide if they want to know any more details." This example demonstrates the tension between the push for quick access on the one hand and doing justice to the complexities of collection histories and contexts on the other.

Locations thus presented many challenges while at the same time proving highly valuable in research (both for mapping species distribution and for provenance purposes; see, e.g. Madruga, 2022; Shavit and Griesemer 2011). To manage the uncertainties the locations presented, I would often ask one of the colleagues working alongside me to crosscheck a label for me. I would pass the label, held between tweezers, over our computer screens to allow the other person to study it. In particularly challenging cases, the scan operator and digitization manager would gather around my computer and puzzle over a location or genus, mumbling potential town names out loud while passing the label around.

When none of us could figure a name out, we would pass more problematic labels on to the curator, who would take out her magnifying glass in a final attempt at documenting the correct information. However, always having to consult colleagues was time-consuming. It therefore became common practice in ambiguous cases to add a small slip of paper with a question mark to the lot (Figure 5). This allowed the inventorizer to indicate any uncertainties or missing information regarding a certain lot, hoping that the scan operator might be able to figure it out during the scanning process. Thus, as part of the digitization process, we found ourselves adding more layers of paper (newly printed inventory numbers, QR codes, question marks) to the lots, further sedimenting labeling practices.

The fact that researching locations is a time-consuming and absorbing process also became clear to me when I visited a malacology collection in the southwest United States, where I met two volunteers, 74 and 84 years old respectively, responsible for taking care of and digitizing the malacology collection. I asked the older of the two what had prompted her to volunteer at the collection: "I'm 84, and I got sick of watching television. Here I get to look at lots of variation." She lifted her hands. "I'll do

it till my hands cramp, which is about two hours. Then I'm out." When I asked the volunteers how the process was going, her younger colleague told me, "We are digitizing, but we are behind. I doubt we'll ever get there … work is very slow, especially figuring out the locations. Just today, I spent an hour looking for a location in Hawaii." Jokingly, her colleague added, "We are making snail progress here." In addition, for these digitizers in the US, the work brought back memories of past travels: "A lot of the places on the labels I have been to, and they bring back memories for me." Digitization was a project that allowed them to step out of their daily routines—to drift off to other places and times.

While digitization is a future-oriented project aimed at offering solutions to biodiversity loss, the section above demonstrates that it requires inventorizers to grapple with the very historicities of natural history collections: through their digitization work, inventorizers come to reconstruct the times and places of these specimens' collection contexts. For the collection of Brandenburg snails, this meant, for example, tracing the shifting boundaries of the German Empire, which had extended into parts of contemporary Poland and Russia in the late nineteenth and early twentieth centuries, and encountering the nationalist politics of script. Or in the case of our North-American volunteers, it meant bringing back memories of past trips. This shows that digitization involves the imaginative work of recollection: both of natural history as a geopolitical project and of the digitizers' own lives and connections. As the historian Catarina Madruga puts it, "The metadata that accompanies collection objects and clarifies where, when, and how specimens were collected in the field is at the core of research on colonial, labor, and environmental history" (2022: 9). However, scaling up the processing of specimens risks losing the historical richness conveyed by these labels. Because the descriptions and stories on the labels were not deemed immediately relevant to scientific biodiversity-monitoring projects and did not contribute to the goal of digitizing specimens quickly, they were not transcribed into the data tool.

Platzprobleme: Mass-digitization and expansion

At the time I was conducting my fieldwork, my colleagues were still experimenting with the development of standards for transcribing labels and working out the optimal DORA settings for different kinds of shells. Our work was not exactly streamlined or smooth yet; we had not yet reached the "mass" scale, which is what in part allowed us to linger over difficult handwriting. What was being anticipated, however, was the eventual "scaling-up" of data production, for which the museum had received funding in 2018.[8] The snailforce was to be expanded with the addition of seven workers and three additional DORAs, making it possible to process specimens faster.

In the following section, I will draw attention to this forward-looking aspect of mass-digitization, arguing with Adams et al. (2009) that it works through logics of anticipation and expansion.

Natural history museums have long been characterized as "centers of calculation" and accumulation (Latour, 1987: 225), making space a recurring concern. As the historian Bruno Strasser notes in his work on "collecting nature," the "collecting sciences" face the challenge of "bring[ing] spatially dispersed objects to a central location and mak[ing] them commensurable." "Collecting," Strasser writes, "was (and is), above all, a spatial practice" (2012: 313). Indeed, due to the physical limitations on storage room in the MfN's main building, collections inevitably presented "Platzprobleme" [space problems], as a colleague in the mammal collection puts it—and not just in terms of collection space. Looking back at the history and context within which the term *biodiversity* emerged in the 1980s, finding the space to count and document species has always been a central concern. As Edward Wilson (1985: 701) wrote, "I believe that we should aim at nothing less than a full count, a complete catalog of life on Earth. To attempt an absolute measure of diversity is a mission worthy of the best effort of science." In the same publication, he raised "information storage" as a quantitative issue, a math problem, writing out a calculation for the meters of shelf space required to document all living species on paper (Wilson, 1985: 703).

The digitization of collections was presented by some as a solution to such spatial issues: it was speculated that digitized collections might make it possible to store physical specimens in a "cheap facility far away from the city center."[9] But mass-digitization is making demands on space in novel ways (Vollmar et al., 2010), forcing digitization managers, for example, to consider picture quality: high definition images require the "stacking" of 20 to 100 photos per specimen. Besides this taking up "a lot of computer power and a lot of time," as one museum photographer explained, it "takes up too much space digitally." Of course, the server space required to store these data must in addition be maintained and paid for, and data storage is not without its own environmental impacts (Monserrate, 2022).

To help us consider the expansiveness of mass-digitization further, we can turn to the work of Adams et al. (2009), who in their piece on anticipatory regimes discuss spatiality through the notion of expansion:

> Anticipatory regimes, like those of capitalism, tend to work through logics of expansion, in which new territories for speculation must be continually found to keep the anticipatory logic moving. Anticipatory regimes expand their scope of inclusion, elongate their reach in time, in space, and in phenomenological terms. (Adams et al., 2009: 250–251)

As Adams and her colleagues put it, anticipatory regimes "work outwardly" (2009: 251), extending temporally and spatially in different ways. The process of mass-digitization works by a similar expansive logic. Indeed, in anticipation of scaling up digitization, anxieties about collection space started to grow. These included concerns about where to place the scanners as well as the desks for the new workers. But there were also more mundane concerns: where were the new colleagues going to have lunch? In which fridge would they put their snacks? Which bathroom would they use?

These new workers would not be employed by the museum but would be hired by an external contractor. They were going to work on short-term contracts, in shifts, inventorizing or scanning a set number of lots per shift. Because they would not be hired by the museum, the workers were not going to have access to museum-wide meetings, workshops, reading groups, or language courses. They were not meant to stay at the institution long-term. This was a shame, my malacology colleagues stressed, because it was valuable to have people on board who had been through different stages of the project, who had seen mistakes being made, and who had acquired the expertise to fix problems—who knew what works and what does not. In this sense, mass-digitization functions on the precarious timescales of capitalist production. Furthermore, the scaling-up and formalization of digitization work reduces the space available to linger over labels: the focus shifts from the more investigatory, collective, and playful work of label transcription toward productionism. The digitization process becomes "subordinated to the linear achievement of future output" (Puig de la Bellacasa, 2015: 706). Performing transcription work remotely on the basis of digital images, away from the collection, was part of this imperative. The digitization manager told me that "having the snails in front of you" slowed down the process of digitization. Being physically present in the collections, people became "too invested" in transcribing the labels correctly; it became hard to "let it go," as she puts it. By being among the snails, in other words, one ran the risk of drifting too far away.

Mass-digitization can be considered expansive in that it draws in an ever-larger host of human and non-human actors and materials. It not only demands more workers and novel arrangements of logistical processes; rather, the mass-digitization of specimens requires paper QR codes, hard drives to store files and digital images, costly server space, plastic containers, and technologies such as hand scanners, tweezers, and scissors. To make room for the new plastic containers used to rehouse the shells, which came in all different shapes and sizes to accommodate the widely diverse shells, a whole section of the collection cabinets, for example, had to be cleared out. Within the digitization projects, those managing the work and timelines would moreover frequently refer to ways in which the data could

be further expanded upon at some point in the future. For example, while location and collector information was only being recorded as "minimal data" while I was conducting my fieldwork, my colleagues told me they would be adding more detailed data "later on." As Akhil Gupta writes, "infrastructure is almost always built to exceed present needs: it is built in anticipation of a not-yet-achieved future" (2018: 63). Furthermore, collections are not static: new specimens continue to be collected and added, and new kinds of information, such as genetic samples, are being connected to existing kinds. In this sense, the project of digitization (and natural history more generally) is without end, never really complete; or, in the words of Adams et al. (2009: 256), "As the mode of anticipation does not need actual objects or events, but must only imagine them as possible, the scope of optimization is unlimited."

Finally, more broadly speaking, space is relevant to the mass-digitization of collections because it reminds us that natural history has been part and parcel of a territorializing project of empire (Madruga, 2019; Nadim, 2021). Natural history collecting, "like establishing empires, requires a mastery of space" (Strasser, 2012: 313). As noted by Strasser, natural history collecting "produced a movement of natural things, which were often dispersed across the world, toward central locations, just as empires produced movements of goods from colonies to metropoles" (313; see also Barringer and Flynn, 1998). Most of the MfN's collections, and therefore the specimens that are currently being digitized, were collected outside Germany's borders, especially in its former colonies (Heumann et al., 2018; Kaiser, 2023). With mass-digitization, the aim is to make these collections accessible globally. However, the reality is that the valuable physical (type) specimens remain in the old imperial centers, and it is there that decisions are made about *what* data is recorded and *how* it is distributed—indeed, where "the movement of natural things" is managed. As such, the project of mass-digitization may be viewed as "the sociomaterial terrain," where colonial systems of power and the unequal distribution of resources are being in part reproduced (Appel et al., 2018: 2; see also Kaiser et al., 2023). This becomes all the more poignant when we recall the purpose of digital collection data—namely, to fuel research into climate change and species loss—and consider the geographical distribution of the drivers of global warming on the one hand and the areas and peoples most affected by it on the other (Pulido, 2018).

Drifting off

This paper has sought to add to understandings of the politics of mass-digitization (Thylstrup, 2018) by zooming in on practices of data production and the way that different temporalities surface within them. I started out by delineating how concerns about pace (speed in particular) appear to

dominate discussions surrounding mass-digitization, with digitization being presented as a linear, easily scalable project, and the resulting data framed as a solution to biodiversity crises. Such an understanding of mass-digitization leaves little space for engagement with the social, historical, and material complexities of collections, making it tempting to argue for more "slowness," especially considering the figure of the snail. As Puig de la Bellacasa, however, points out in her work on the pace of soil care, slowness does not offer a way "out" of the linear, progressivist notion of time as speed, entangled with innovation and new technologies: "Advocating slowness as time of a different quality against the speed of innovation and growth in technoscience does not necessarily question the direction of the dominant timeline, which these approaches do by operating differently within technoscience" (Puig de la Bellacasa, 2015: 709). Slowness, even if it may operate differently, is still a form of speed: it qualifies the movement of an object over distance, implying a progressive, forward-looking trajectory. From snails we might instead learn the art of *drifting*, forgoing direction and control to allow for improbable journeys.

Snails' capacity to drift, field philosopher Thom van Dooren argues in his work on Hawaiian terrestrial snails, can be characterized as a form of deep time travel. Floating and flying around the world's oceans, snails, without human intervention, have made it to islands all around the world: "At some point in the distant past, a tiny snail climbed on board a migratory bird, perhaps a golden plover, as it perched or nested overnight … Days or weeks later, having rested through the exhausting crossing, the snail then climbed off the bird in its new home." While these journeys, as Van Dooren recognizes, are highly unlikely, they lead to places that matter. My main takeaway is that they demonstrate the potential affordances of a lack of directionality, how it opens up unexpected destinations. In her analysis of drift wood (incidentally, another vehicle that snails have used to float around on) washing up and accumulating on the shores of Norway and Iceland, Þóra Pétursdóttir has likewise used the notion of drift, of veering off course, to rethink the meaning of heritage and its management. Telling stories of waste, discarding, and neglect, she points to the potential of drift to help us embrace indeterminacy and change instead of approaching objects of heritage as "compliant subjects of stewardship, management, and control" (2020: 97). Moreover, her deliberations on drifting have inspired others to draw attention to the "drift time" of heritage, characterized as the *mixing* of different timelines (DeSilvey et al., 2020). Drifting, as Pétursdóttir and DeSilvey et al. agree, opens up "alternative ways of knowing and reasoning with heritage matters" (DeSilvey et al., 2020: 457).

Just as objects of heritage themselves are not "compliant subjects of control," neither are the data that result from them. This paper has demonstrated that the practices and processes that give shape to biodiversity data require a continuous engagement and grappling with different temporalities. Attending to these different temporalities in practice—when they surface or become obscured in workflows and inventory or scanning work, when and how they sediment on and around specimens, and when they become (mis-)aligned in work processes—the politics of mass-digitization become tangible. It then becomes clear that output and access to data are being valued over the fostering of skilled workers and their expertise, that the study of biodiversity through taxonomy is being prioritized over investigations into the museum's colonial and extractive histories despite their deep interconnectedness (Raja, 2022), and that there is little critical reflection on the impacts and long-term management of mass-produced data. Whereas these concerns currently play a minor role in discussions around the use and value of the mass-digitization of natural history collections and biodiversity data, they remind us that biodiversity loss is not simply a natural phenomenon, but highly political in itself, resulting from the complex interweaving of historical and socioeconomic processes, including the continuing legacies of colonialism. To adequately address biodiversity crises with data, this complexity should be reflected back in digitization processes and given equal weight alongside more technical considerations. Drifting, then, not only offers alternative ways to engage with objects of heritage but also opens the floor for discussions of how these objects and their data might come to themselves *inform* management.

ORCID iD

Roos Hopman https://orcid.org/0000-0003-3074-2956

Notes

1. A name coined by the engineers responsible for installing it in one of the museum's exhibition rooms.
2. https://www.museumfuernaturkunde.berlin/en/press/press-releases/mass-digitization-insects-world accessed 20/09/2023.
3. Ibid.
4. https://www.nytimes.com/2023/03/23/science/science-museums-online-collections.html accessed 28/03/2023.
5. Interview 5 November 2021.
6. This is true of natural history museums in the US and Canada as well; see Sierwald et al. 2018.
7. As a colleague put it in a presentation on their digitization project, April 20, 2022.
8. https://www.museumfuernaturkunde.berlin/en/press/press-releases/a-new-era-nature-bundestag-approves-660-million-euro-development-berlins
9. Fieldnotes 17 September 2021.

References

Adams V, Murphy M and Clarke AE (2009) Anticipation: Technoscience, life, affect, temporality. *Subjectivity* 28: 246–265.

Ames S and Lewis S (2020) Disrupting the library: Digital scholarship and Big Data at the National Library of Scotland. *Big Data & Society* 7(2): 1–7.

Appel H, Anand N and Gupta A (2018) Temporality, politics, and the promise of infrastructure. In: Anand N, Gupta A and Appel H (eds) *The Promise of Infrastructure*. Durham and London: Duke University Press, 1–38.

Barringer TJ and Flynn T (eds) (1998) *Colonialism and the Object: Empire, Material Culture, and the Museum*. London: Routledge.

Bass MA, Goldgar A, Grootenboer H, et al. (2021) *Conchophilia. Shells, Art, and Curiosity in Early Modern Europe*. Princeton and Oxford: Princeton University Press.

Bates J, Lin YW and Goodale P (2016) Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society* 3(2): 1–12.

Besky S and Blanchette A (eds) (2019) *How nature works: Rethinking labor on a troubled planet*. Albuquerque: University of New Mexico Press.

Bonacchi C (2021) Heritage transformations. *Big Data & Society* 8(2): 1–4.

Boumans M and Leonelli S (2020) From dirty data to tidy facts: Clustering practices in plant phenomics and business cycle analysis. In: Leonelli S and Tempini N (eds) *Data Journeys in the Sciences*. Cham: Springer, 79–101.

Bowker GC (2000) Biodiversity datadiversity. *Social Studies of Science* 30(5): 643–683.

Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, US: MIT Press.

Callomon P (2019) Standard views for imaging mollusk shells. *American Malacological Society:* http://malacological.org/epubs.html.

Daston L and Galison P (2007) *Objectivity*. New York: Zone Books.

Davies G, Frow E and Leonelli S (2013) Bigger, faster, better? Rhetorics and practices of large-scale research in contemporary bioscience. *BioSocieties* 8: 386–396.

DeSilvey C, Bartolini N, Lyons A (2020) Processing change. In: Harrison et al. (eds) *Heritage Futures: Comparative Approaches to Natural and Cultural Heritage Practices*. London: UCL Press, 446–461.

Gitelman L (ed) (2013) *Raw data is an oxymoron*. Cambridge, US and London: MIT Press.

Gupta A (2018) The future in ruins: Thoughts on the temporality of infrastructure. In: Anand N, Gupta A and Appel H (eds) *The Promise of Infrastructure*. Durham and London: Duke University Press, 62–79.

Hedrick BP, Heberling JM, Meineke EK, et al. (2020) Digitization and the future of natural history collections. *BioScience* 70(3): 243–251.

Heerlien M, Van Leusen J, Schnörr S, et al. (2015) The natural history production line: An industrial approach to the digitization of scientific collections. *Journal on Computing and Cultural Heritage* 8(1): 1–11.

Heumann I, Stoecker H, Tamborini M, et al. (2018) *Dinosaurierfragmente: Zur Geschichte der Tendaguru-Expedition und ihrer Objekte, 1906-2018*. Göttingen: Wallstein Verlag.

Hirsch SL, Ribes D and Inman S (2022) Sedimentary legacy and the disturbing recurrence of the human in long-term ecological research. *Social Studies of Science* 52(4): 561–580.

Jackson SJ, Ribes D, Buyuktur AG, et al. (2011) Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work. *Proceedings of the ACM 2011 conference on Computer supported cooperative work*: 245–254.

Kaiser K (2023) The coloniality of natural history collections. In: Andratschke C, Müller L and Lembke K (eds) *Provenance Research on Collections from Colonial Contexts: Principles, Approaches, Challenges*. Heidelberg: arthistoricum.net-ART-Books, 242–255.

Kaiser K, Heumann I, Nadim T, et al. (2023) Promises of mass digitisation and the colonial realities of natural history collections. *Journal of Natural Science Collections* 11: 13–25.

Kitchin R and McArdle G (2016) What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society* 3(1): 205395171663113.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Leese M and Pollozek S (2023) Not so fast! Data temporalities in law enforcement and border control. *Big Data & Society* 10(1): 1–13.

Leonelli S (2019) What distinguishes data from models? *European Journal for Philosophy of Science* 9(2): 1–27.

Leonelli S (2020) Learning from data journeys. In: Leonelli S and Tempini N (eds) *Data Journeys in the Sciences*. Cham: Springer Nature, 1–24.

Madruga C, 2019. *Taxonomy and Empire. Zoogeographical Research on Portuguese Africa, 1862-1881*. PhD Thesis, Universidade de Lisboa, Portugal.

Madruga C (2022) "Authentic provenance." Locality and colonial collecting for the Lisbon Zoological Museum, 1860s–1880s. *Journal for the History of Knowledge* 3(1): 1–13.

Magurran AE and McGill BJ (2011) *Biological diversity: Frontiers in measurement and assessment*. Oxford: Oxford University Press.

Marshall CR, Finnegan S, Clites EC, et al. (2018) Quantifying the dark data in museum fossil collections as palaeontology undergoes a second digital revolution. *Biology Letters* 14(9): 1–4.

Mol A (2020) Not quite clean: Trailing schoon and its resonances. *The Sociological Review* 68(2): 385–400.

Monserrate SG (2022) The cloud is material: On the environmental impacts of computation and data storage. *MIT Case Studies in Social and Ethical Responsibilities of Computing*, no. Winter 2022 (January).

Nadim T (2020) System box (tray) with wasp. In: Bauer S, Schlünder M and Rentetzi M (eds) *Boxes: A Field Guide*. Manchester: Mattering Press, 109–123.

Nadim T (2021) The datafication of nature: Data formations and new scales in natural history. *Journal of the Royal Anthropological Institute* 27(S1): 62–75.

Nadim T, Vennen M, Heumann I, et al. (2024) Logistical natures: Trade, traffics, and transformations in natural history collecting. *Historical Studies in the Natural Sciences* 54(2): 125–134.

Nelson G and Ellis S (2018) The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B* 374(1763): 1–9.

Parry R (2007) *Recoding the Museum. Digital Heritage and the Technologies of Change*. London: Routledge.

Pétursdóttir Þ (2020) Anticipated futures? *Knowing the Heritage of Drift Matter. International Journal of Heritage Studies* 26(1): 87–103.

Plantin JC (2021) The data archive as factory: Alienation and resistance of data processors. *Big Data & Society* 8(1): 205395172110075.

Puig De La Bellacasa M (2015) Making time for soil: Technoscientific futurity and the pace of care. *Social Studies of Science* 45(5): 691–716.

Pulido L (2018) Racism and the anthropocene. In: Mitman G, Armiero M and Emmett RS (eds) *Future Remains: A Cabinet of Curiosities for the Anthropocene*. Chicago: University of Chicago Press, 116–128.

Raja NB (2022) Colonialism shaped today's biodiversity. *Nature Ecology & Evolution* 6: 1597–1598.

Roth WM and Bowen GM (1999) Digitizing lizards: The topology of vision in ecological fieldwork. *Social Studies of Science* 29(5): 719–764.

Schopp JF (2016) "… für das deutsche Wort nur die deutsche Schrift!" Typographie als Zeichen nationaler Identität? In: Barstad GE, Hjelde A, Kvam S, Parianou A and Todd J (eds) *Language and Nation: Crossroads and Connections*. Münster and New York: Waxmann, 265–296.

Shavit A and Griesemer J (2011) Transforming objects into data: How minute technicalities of recording "Species location" entrench a basic challenge for biodiversity. In: Carrier M and Nordmann A (eds) *Science in the Context of Application. Boston Studies in the Philosophy of Science, vol 274*. Dordrecht: Springer, 169–193.

Sierwald P, Bieler R, Shea EK, et al. (2018) Mobilizing mollusks: Status update on mollusk collections in the U.S.A. and Canada. *American Malacological Bulletin* 36(2): 177–214.

Star SL and Ruhleder K (1996) Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *Information Systems Research* 7(1): 111–134.

Star SL and Strauss A (1999) Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work* 8(1): 9–30.

Strasser B (2012) Collecting nature: Practices, styles, and narratives. *Osiris* 27(1): 303–340.

Thylstrup NB (2018) *The Politics of Mass Digitization*. Cambridge, MA and London: The MIT Press.

Van Dooren T (2022) *The World in a Shell. Snail Stories for a Time of Extinctions*. Cambridge, US: MIT Press.

Vennen M (2022) Keeping Records of Animals. In: Heumann I and Nadim T (eds) *Animals as Objects?*. https://animalsasobjects.org.

Vollmar A, Macklin JA and Ford L (2010) Natural history specimen digitization: challenges and concerns. *Biodiversity informatics* 7(2): 93–112.

Whitehead PJP (1971) Storage and retrieval of information in systematic zoology. *Biological Journal of the Linnean Society* 3(3): 211–220.

Wilson EO (1985) The biological diversity crisis. *BioScience* 35(11): 700–706.

Wilson MW (2011) Data matter(s): Legitimacy, coding, and qualifications-of-life. *Environment and Planning D: Society and Space* 29(5): 857–872.