Herbaria Heritage: Visualizing Colonial Bias in Natural History Collections

Sakura Morales Furuta[1*], Ana Reviejo Salamanca[2*], Michaela Todd[1*], Lise Stork[3], and Andreas Weber[4]

[1]Leiden University, Leiden, The Netherlands
[2]Vrije Universiteit, Amsterdam, The Netherlands
[3]University of Amsterdam, Amsterdam, The Netherlands
[4]University of Twente, Enschede, The Netherlands

As a result of the colonial entanglements of many natural history collections, much of the world's biodiversity heritage is housed in Europe. Increasingly, natural history institutions have begun to address their colonial past. However, in this regard, computational methods for analyzing large collections tend to consist of static visualizations of collection provenance. Moreover, in regions where cutting-edge visualization technologies are only scarcely available - often these are the same areas from which most European plant collections originate - there is a lack of simple, accessible solutions that can generate meaningful results. Thus, we argue that accessible, simple, yet interactive visualizations of collection provenance allow users to understand colonial bias in natural history collections better. Our solution allows users to focus on content gaps and highlights historical patterns and trends in collection data. Using a dataset containing metadata of five million entries from the Naturalis Biodiversity Center botanical collection as a use case, we created an interactive visualization with Microsoft Power BI. The visualization showcases the origins and movements of botanical specimens from former Dutch colonies to the Netherlands on an interactive map and timeline. This addresses not only a gap in historical research on the colonial legacy of Dutch botanical collections but also a gap in digital humanities research regarding simple and easy-to-use computational techniques for distant reading of natural heritage data. The particular use case demonstrates only a fraction of the research possibilities that this tool enables. Our interactive visualization increases the accessibility of the available scientific data, and contributes to a better understanding of the relationship between cultural history and natural history, highlighting

* Sakura Morales Furuta, Ana Reviejo Salamanca and Michaela Todd share the first authorship of this paper.

the greater importance of easy-to-use, interactive and accessible visualizations of biodiversity collection histories. Ultimately, our project suggests a way forward for natural history museums not only in the Netherlands to reinterpret the colonial past of their collections.

Keywords: visualization, colonial bias, natural history collections, digitization

1 Introduction

While Natural History museums and herbaria in the Global North house much of the world's biodiversity heritage, they have tended to disassociate themselves from the cultural histories inherent to their collections (Borren and Drieënhuizen, 2022; Das and Lowe, 2018; Drew et al., 2017; Kaiser et al., 2023). This lack of recognition of the histories behind collections becomes even clearer when situated alongside other museum disciplines that have been grappling with this problem for years (Haarnack, 2021). However, acknowledging the inherent colonial entanglements of natural history collections is the first step in writing more nuanced histories of these collections. This entails a greater analytical focus on the histories of acquisition and appropriation in the field (Dubald and Madruga, 2022; Hünniger, 2021) and on the subsequent movement of objects to Europe. Such a focus ensures the political context of collections is not hidden under the guise of biodiversity science and rationality, as is often the case today (Driver et al., 2021; Nadim et al., 2024). Additionally, this type of recognition becomes vital for understanding the implications of such colonial entanglements for global biodiversity and conservation efforts (Johnson et al., 2023; Park et al., 2023).

The work done to uncover these hidden histories has tended to concern researchers working primarily with textual or object-focused approaches, lacking computational methods and visualizations (e.g. Alcantara-Rodriguez et al. (2022); Drieënhuizen and Sysling (2021); Jacobs and Koch (2021); Müller (2023); Wingerden (2020)). Some studies by Mohammed et al. (2022) and Park et al. (2023) utilized heat maps, world maps, etc. to illustrate botanical specimen movements across the globe and categorize trends. However, these existing visualizations, such as herbaria snapshots, lack interactivity, leaving a rich trove of information untapped. This lack of interactivity means users cannot currently engage with the different temporalities that natural historical collections and their digitized counterparts entail (Hopman, 2024). Moreover, in localities where advanced and cutting-edge visualization technologies are limited – often the same regions where many European plant collections originate – there is a shortage of simple, accessible tools capable of generating meaningful results (Buschke et al., 2023). Our tool is therefore not intended to be cutting-edge from a technological point of view. On the contrary, we seek to provide a simple, accessible and user-friendly solution that can be used as widely as possible.

Given the increasing digitization of natural history collections and the availability of large datasets related to these collections (Beltrame et al., 2024; Hedrick et al., 2020; Pickering, 2024; Stork et al., 2019, 2021), there is great potential to exploit the data in *simple*, *meaningful* and importantly, *interactive* ways. Thus, we propose *interactive visualizations* as a computational technique for distant reading of natural heritage data. Focusing on the provenance of botanical specimens – specifically those stored in Naturalis Biodiversity Center (Leiden, The Netherlands) – we create an interactive and multi-layered visualization to trace historical movements of species from Indonesia

and Papua New Guinea[1] using Microsoft Power BI (see Figure 2). Our dashboard and accompanying use case are insightful for the processing, design and visualizations of other collection data. Since Microsoft BI is already used in the domain to visualize digitization progress and the composition of collections, we foresee a fast uptake of our solution (Breugelmans and Trekels, 2023; DiSSCo, 2023).

We are of course aware that there is a deep disciplinary and institutional divide between the natural sciences and cultural, colonial and historical studies. While natural science oriented researchers consider digitized natural history collections merely as abstract and decontextualized datapoints, historians read them as entangled cultural objects with a long history of collection, appropriation, circulation and datafication. While some have suggested that hiring more diverse staff at natural history museums is the way forward to achieve a more holistic understanding of such collections (Das and Lowe, 2018), we think that an intensified training of historians in critical and reflective digital humanities, for instance with a focus on interactive visualization techniques, should play an important role to fill this epistemological and methodological gap (Díez Díaz et al., 2025). A smart use of interactive and easy to use visualization tools such as Power BI can help to reconnect datasets to their cultural and historical roots. Taken together, we hope that our paper serves as inspiration for (biodiversity) scientists, historians, data providers and a general public - not only in the Western world - to actively continue and engage in conversations about new ways of access and visualization of collection datasets in the field of natural history.

To outline our research, we first describe the dataset (Section 2), before detailing our use case and the historical, practical relevance of the case. Secondly, we compare existing computational methods to further highlight the gap we aim to address in the related work (Section 3). With this background, we describe our method (Section 4), before discussing the obtained results from a qualitative perspective (Section 5). In the final section (Section 6), we reflect upon our methodology, suggesting follow-up work. Ultimately, this article aims to bridge the gap between the historical and digital, exploring how data visualization can be used as a tool to uncover the histories hidden behind scientific collections.

2 Dataset

Our paper employs the "Naturalis Biodiversity Center (NL) - Botany" dataset, obtained through the GBIF website (Creuwels, 2015) [2]. GBIF, or the Global Biodiversity Information Facility (GBIF), is an international network and data portal that aggregates data from natural historical institutions all over the world (Feng et al., 2022). The physical plant specimens that this dataset represents are stored in the depots of Museum Naturalis in Leiden. This freely available tabular dataset comprises more than five million entries registering botanical specimens found in the Naturalis natural history collection. It has 51 columns with metadata about collected specimens, from which we use: country of origin (referred to as *country* in the dataset), *recordedBy, continent, family* and *typeStatus*. These five categories have been implemented as filters

[1] We are aware of the complex colonial history of Papua New Guinea. In-depth historical research would be necessary to find out which - and even more important when - specific botanical specimens were registered under the locality of Papua New Guinea.

[2] The dataset can be downloaded through the following link: `https://www.gbif.org/dataset/15f819bd-6612-4447-854b-14d12ee1022d`, however, its download is not necessary in order to run the Power BI file.

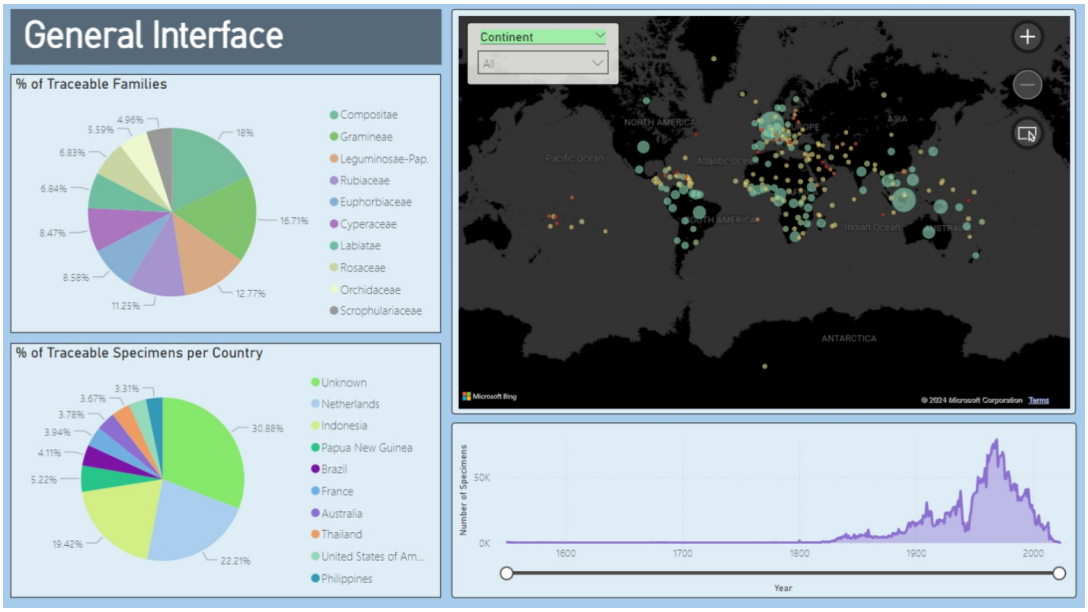

Figure 1: General Interface, including two pie charts, map, and timeline.

within the interactive interface, as well as through a world map to plot specimen occurrences.

Looking more closely at the data from Figure 1, the Netherlands accounts for 650,464 (22.21%) of Naturalis' specimens, followed by Indonesia at 568,517 (19.42%) and Papua New Guinea at 152,914 (5.22%). Taking the *Orchidaceae* family as an example (see Figure 3), the visualization reveals that Indonesia leads with 19,695 occurrences (37.38%), followed by 'Unknown', with 13,519 specimens (25.66%), and finally Papua New Guinea with 8,437 occurrences (16.01%). Significant spikes in registered specimens occurred in 1909 (907 from 51 countries), 1936 (1,328 from 34 countries), and 1986 (1,791 from 47 countries). Orchids are interesting, since in the nineteenth and early twentieth centuries, orchids from colonial areas were an important economic luxury good and played an important role in scientific debates within the botanical sciences. Not only for individual botanists, but also for scientific institutions such as herbaria or natural history museums, it was important to acquire and publish about new specimens collected in colonized areas (Endersby, 2016, pp.105-128). Next to orchid hunters, colonial botanical gardens such as the one in Buitenzorg (now Bogor, Indonesia) or gardens in British India played key role in establishing global networks of plant exchange (Baber, 2016; Drayton, 2000; Goss, 2011; Weber and Wille, 2018).

3 Related Work

3.1 Colonial Roots of Natural History Collections

Although a contentious issue, an important body of literature exists exploring various topics which relate to the broader issue of addressing the colonial roots of natural history collections (Coote et al., 2017; Curry et al., 2018). There is consensus surrounding the finding that it is not uncommon for natural history museums to present their collections in a way that blurs "its entanglement with the human world" (Drieënhuizen and Sysling, 2021, p.305). For instance, scholars have found that Eurocentric perspectives

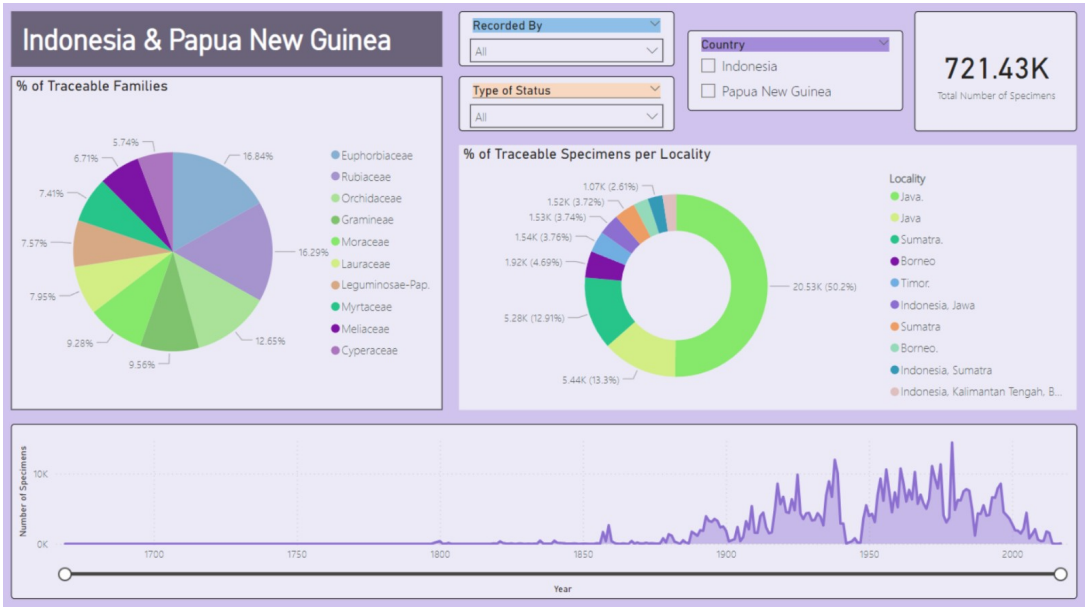

Figure 2: Indonesia and Papua New Guinea Interface, including a pie chart, filters, a donut chart and a timeline.

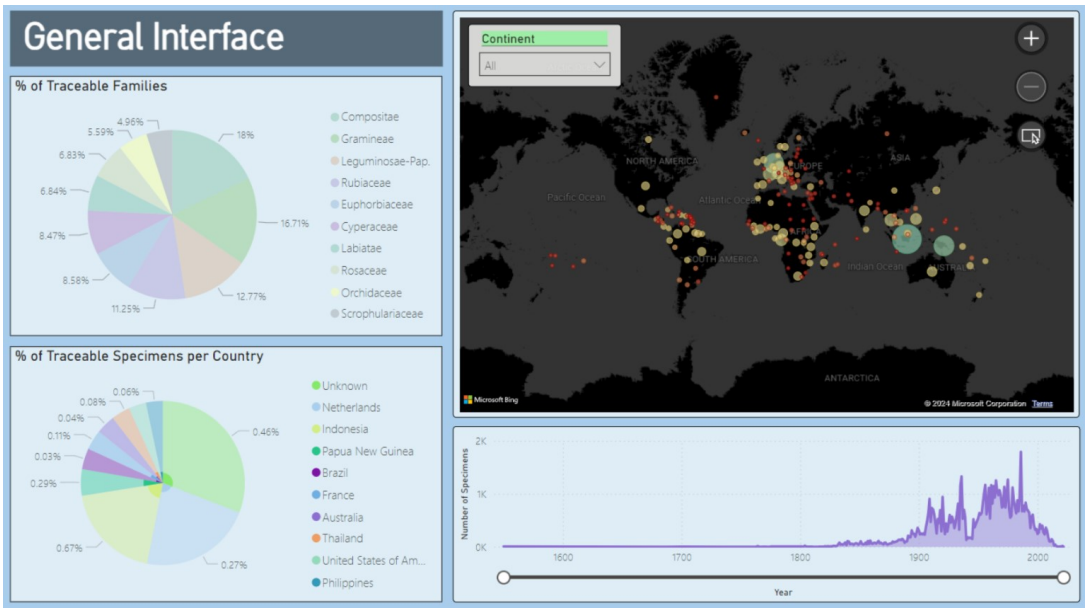

Figure 3: General Interface with the Orchidaceae family selected, shifting the visualizations to be filtered by this family.

in natural history museums have neglected indigenous contributions, perpetuating troubled social histories and emphasizing the need for decolonization (Andel, 2017; Ashby and Machin, 2021; Leonard, 2024). Mohammed et al. (2022), for example, reveal how colonialism erased indigenous knowledge in the Caribbean archipelago, creating biases in biodiversity collections. Similarly, the 'coloniality of power' concept coined by Quijano (2007), as discussed by Kaiser et al. (2023), highlights ongoing epistemic forms of extraction and racism during post-colonial rule, influencing scientific practices and collection growth. From those working within the natural sciences field, attempts to address such colonial roots have recently received more urgency, with organisations such as NatSCA (Natural Sciences Collections Association) hosting conferences to tackle this issue.[3]

Interestingly, however, (Park et al., 2023, p. 1060) note that the focus has largely been on "human- and animal-related collections" with comparatively less attention paid to botanical collections, possibly due to "lower visibility; few herbaria offer public displays, and plant awareness is generally lacking". Moreover, there are only a few works which investigate the parallels between specimen acquisition and historical events from a longer term historical perspective. For example, researchers such as Goss (2011, 2018, 2023); van de Loosdrecht et al. (2024); Weber and Wille (2018); Wille (2019) uncover a connection between botanical research and the Dutch colonial project in Indonesia from a historical perspective. Nevertheless, work that uncovers the colonial entanglements of natural history collections of former colonizing countries – especially botanical collections – through the lens of large datasets remains limited and thus presents a gap to which our work aims to contribute.

3.2 Visualizing Colonial Bias in Botanical Datasets

Computational approaches addressing the colonial roots of natural history collections are limited, particularly concerning visualization tools. Looking more broadly at visualization techniques, several options exist, albeit more tailored toward cultural heritage – rather than natural history – and largely limited in interactivity. These include Information Visualization (InfoVis), which involves visualization methods such as graphs, charts and maps (Windhager et al., 2018). In a broader exploration of herbarium collections and their colonial legacies, (Mohammed et al., 2022, pp.147-148) employed visualizations to trace specimens from Trinidad and Tobago and the Bahamas. Their use of a heat map displaying institutions and collection sizes, along with a stacked bar chart illustrating accessioning patterns within the University of the West Indies Zoology Museum, effectively highlighted the historical role of islands as natural laboratories and the lasting impact of colonial influences on scientific practices. Related to this, Park et al. (2023) conducted a global investigation, examining herbarium disparities and presenting their findings through various visualizations. For instance, a world map visualized plant specimen movements, while relationship charts depicted intercontinental flows before and after colonialism. Their use of histograms in combination with maps provided a comprehensive illustration of the percentage of internationally collected herbarium specimens and categorizing trends across institutions and continents. These are both examples of how visualization was leveraged as a tool to increase the accessibility of natural history data, uncovering the colonial bias of collections in the process.

However, as noted by (Boakes et al., 2010, p. 1), "For these and related purposes,

[3] See for more details: `https://www.natsca.org/event/2605`, accessed 19 October 2024.

biodiversity information must include more than a simple snapshot of the current status and distribution of species." The static nature of these techniques, including the R package bdvis,[4] is self-evident (Barve and Otegui, 2016; Medeiros e Sá et al., 2022). Multi-Touch interfaces, on the other hand, offer greater interactivity by enabling the intuitive exploration of collections, such as browsing museum image collections on multi-touch tables (Windhager et al., 2018, p.12). However, these are largely focused on exploring collection items rather than the distant reading of data. Moreover, 3D visualizations offer 3D depictions of artifacts or sites, but again, this visualization method is limited in applicability to our specific research dilemma (Windhager et al., 2018, p.15). A similar limitation exists for augmented reality and virtual reality. A technique more relevant for our purposes – specifically, contextualizing collections – is narrative visualization, which guides users through collections using storylines, timelines or spatial encodings. Again, however, interactivity within the visualization is a key element that is often missing. The same can be said for network visualizations and faceted browsing. Thus, more relevant for our purposes are 'generous interfaces', which allow users to interact serendipitously; and interactive timelines which in our case, are particularly important to understand the colonial contexts behind collections (Windhager et al., 2018, pp.1-15). Although interactive visualizations of museum collections do exist, there is a clear gap in the field of natural history collections specifically. This is despite archival and digitized collection databases being more openly available. It has been acknowledged by scholars that "the world's museums contain information that is irreplaceable, especially a unique historical perspective, yet their information is often hard to access, even in the case of some well-funded, national museums" (Boakes et al., 2010, p.7). Thus, there remains a gap when it comes to making natural history museum data more accessible, especially for low-resource settings. Building on the above-mentioned visualization techniques offers a promising solution.

Ultimately, our paper argues that *multilayered, interactive* visualizations which allow users to explore and understand the history and provenance of scientific collections using any openly available data are lacking. This is especially problematic in settings where computational power and resources are limited. Research on the design of interactive health-related data dashboards in low resource contexts can serve as an inspiration here, for instance Akpan et al. (2022); Mgusha et al. (2021), which both utilize Power BI. The unavailability of interactive dashboards in the context of large natural history-related datasets restrains the range of research questions that can be asked: important stories that lie beneath the surface of the data are unlikely to be revealed, and the serendipitous discovery that often prompts further investigation is prematurely stunted.

4 Approach

For our visualization we used Power BI Desktop, which allows interactive maps, timelines and filters based on database column headers. There are different reasons why this specific tool was put to use. First of all, Power BI is an established tool for visualizations of collection data in the field of natural heritage collections (see Collection Digitization Dashboard, `https://www.dissco.eu/services/#cdd`). Moreover, it is an established way to create interactive data visualizations in other scientific disciplines, e.g. health science Akpan et al. (2022); Mgusha et al. (2021). Second, we were

[4] `https://github.com/vijaybarve/bdvis`

looking for a tool that is freely available, enables replicability, and is easy to use. The tool allows for an interactive, user-friendly experience without requiring advanced programming knowledge. This makes replicability accessible, not only to historians and scientists but also to museum professionals, policy makers, and the general public. Thirdly, we were looking for a solution that does not require extensive computational resources and energy. Since the tool maintains its interactive performance even when handling large datasets, we opted for Power BI. We created dashboards on a laptop PC with a Ryzen 5 5500U processor and an AMD Radeon Graphics GPU. Since the selected Dutch botanical dataset from Naturalis is extremely large (with about more than five million data entries), we decided to create two views on the data: a general interface (see Figure 1) for an overview of the database, and a focused one (see Figure 2). For the focused one, we chose a focus on Indonesia and Papua New Guinea due to their colonial history, allowing researchers to get an overview of *when* Dutch researchers registered specimens native to these two countries, to understand *how* the collection's colonial bias has evolved over time and space.

4.1 Preprocessing

The "Naturalis Biodiversity Center (NL) - Botany" dataset was processed by: collapsing date ranges from *eventDate* into single dates (some were single-day timestamps, and some spanned one year), and removing the empty, null and "?" values, which was done through filters within the software. To replace the dataset with another, the same processing steps need to be repeated.

4.2 Solution Approach and Design

After loading the data into Power BI Desktop, the next step was to decide how to best represent it visually, in order to achieve our overarching aim of understanding the colonial bias of this collection. Given the geographical focus of our project, the most appropriate starting point was a map; specifically, one which visualizes the number of specimens in the Naturalis collection that originate from a particular country. Size and color were used to differentiate (e.g. the larger the circle, the more specimens that originate from that country). The interface also shows two pie charts, one with the percentage of traceable families, and another with the percentage of specimens per country. Importantly, there is a timeline plotting the number of specimens collected over time (see bottom of Figures 1 and 2). All visualizations are interactive, increasing the user-friendly nature of the tool. Given that previous computer science knowledge is not a prerequisite to interact with such a tool, the accessibility of the data improves substantially. Moreover, despite the size of the dataset, Power BI Desktop only needed a fraction of the available computational resources to compute the above-described interactive dashboards.

In terms of the design, the interface is aimed at both historians and scientists as well as the general public with a particular interest in this kind of data. Therefore, we prioritized readability and simplicity but also ensured that a more detailed investigation can be conducted where necessary through the use of the aforementioned filters, which increases the interactivity possibilities of the visualizations (top middle on Figures 1 and 2) and more importantly, the research questions that might be answered. Some of the filters, such as *TypeStatus*, might not be familiar to all users of the dashboard. Selecting this filter lets the user classify specimens by their type. An example of type is holotype. A holotype is a physical plant specimen (e.g. a dried plant) that was used for

the formal description of a species (for more details, see Sluys, 2021). The *recordedBy* filter lets the user filter by the botanist who obtained and recorded the specimen for the Naturalis collection. This directly contributes to questions about the colonial links of collections, as many collectors and botanists (e.g. Carl Ludwig Blume, Pieter Willem Korthals, Jaheri, Kees van Steenis) were closely affiliated with the botanical garden in Bogor, Indonesia and 's Rijksmuseum van Natuurlijke Historie (National Museum of Natural History), as the institutional predecessor of the Naturalis museum in Leiden was called (Goss, 2023; Jacobs and Shea, 2022; Manse, 2013; Miracle, 2021; Sysling et al., 2024; Weber, 2018; Wingerden, 2025). The capability of creating multiple interfaces with the same database illustrates the future possibility to broaden the scope of the relatively small-scale project as it is described in this paper.

5 Qualitative Evaluation

The following section evaluates the functionality of the interface, highlighting its ability to display and filter data accurately while facilitating accessibility without requiring the original dataset. The file is easily shareable, and a receiver does not need the dataset (unless they aim to make changes to the underlying data) in order to access the visualization. The interface is responsive, fast and accurately displays the selected data when run on the aforementioned device. This is key for improving the overall accessibility of such large datasets. Furthermore, the chosen filters (such as the *Orchidaceae* family filter on the second interface) function perfectly and can be modified at any time. Through the malleability of the visualizations, many historical or scientific questions relating to the colonial bias of natural history collections can be addressed. Nonetheless, it would be necessary to obtain a larger dataset with more than only botanical specimens to fully uncover the colonial bias of natural history museums. Given the fast-growing number of datasets of Dutch and other natural historical collections, we see ample opportunity for future research in this field (García-Roselló et al., 2023).

When evaluating the interface from a qualitative point of view, we can conclude that it showcases historical trade and collection patterns of various plant species, notably in Indonesia and Papua New Guinea, while linking specimen collection trends to significant colonial and scientific events, demonstrating the potential for deeper exploration of colonial biases in natural history collections. Looking at the map on the first interface as well as the pie chart, there is a significant clustering of specimens obtained from South-East Asia, most notably Indonesia. It comprises 19,42% of the total dataset, when not accounting for occurrences which are 'Unknown', or 'Null'. More importantly, according to the dataset, 70,23% of all holotype specimens in the Naturalis plant collection (=14.580) stem from Indonesia, Papua New Guinea and Suriname. The unknown cases also remain relevant: (i) they provide museums with a starting point for improving their datasets by highlighting data gaps, and (ii) could be a signal: such specimens could have come from former Dutch colonies, or perhaps locations that museums prefer not to publicize. The unknown cases in our dataset remain, therefore, relevant as they provide museums with a starting point for improving their datasets by addressing data gaps that biodiversity researchers have not considered before. While biodiversity researchers usually focus on specimens and inconsistent metadata (e.g. Boakes et al. (2010); Franz and Sterner (2018); Hughes et al. (2024); Maldonado et al. (2015)), our approach offers a tailored solution to contextualize these gaps within historical and colonial frameworks.

There are several mechanisms to expand approaches for addressing the aforementioned data gaps in collection-related biodiversity research. This involves improving metadata consistency, enhancing specimen traceability, and integrating standardized data-sharing practices (Franz and Sterner, 2018; Hughes et al., 2024; Maldonado et al., 2015). However, beyond these technical solutions, there is a need to contextualize such gaps within historical and colonial frameworks to understand their broader implications. By engaging with biodiversity researchers and data providers, future research in the field of data visualization could work toward refining datasets in ways that not only enhance their scientific utility but also uncover hidden biases in data collection and representation (Islam et al., 2022; Weber and Turnhout, 2024). Incorporating colonial histories into data improvement strategies may help museums and researchers acknowledge the socio-political forces that shaped natural history collections. Addressing these gaps through interdisciplinary collaboration with a specific focus on visualization allows for a more holistic understanding of both the scientific and historical significance of these collections (Viola, 2023).

As mentioned, the dates on which each specimen was collected were visualized in a simple interactive timeline. Peaks and declines in registered specimens can easily be detected this way, thus allowing for parallel historical research. The peaks in 1918 and 1920 might be associated with the onset of the idea of 'pure science', which was used to portray the Dutch colonial project as "running a moral and decent colony, generating scientific knowledge untainted by conquest or occupation ideology" (Goss, 2009, p.191). Another peak occurred in 1970 - 1979, which can be explained by the rapid advancements in botany, particularly in orchid biology. Other peaks, though smaller, are distinguishable, such as in 1936 and slightly in 2004. These provide impetus for future research. Drops in the timeline also correlate to specific historical events. In the 1800s, only 1 or 2 specimens were recorded until the 1890s, possibly due to the lack of documentation or interest in the Orchidaceae family of plants. The fall in 1914 - 1915 can be attributed to World War I (Hiepko, 1987). With World War II alongside the invasion of the Dutch East Indies by the Japanese, rates of occurrences from Indonesia drastically decreased, from 237 specimens recorded in 1939, to 1 recorded in 1945 (Boomgaard, 2008). Other lows can be attributed to other events, such as anti-western and anti-imperial ideologies in the Indonesian government during the 1960-1970s (Boomgaard, 2008).

Moreover, the general interface also allows us to zoom in on specific plant families (Figure 4). Among the species shown in the percentage of traceable families, commercial and traded specimens can be identified. The numbers explained below offer historians and other researchers concrete points of departure for their contextualization of the growth of specimens belonging to specific plant families. Approaching the evolution of the collection's colonial bias through the lens of economically valuable plant families is another area in which our dashboard yields new visual insights. *Rubiaceae*, for instance, most commonly known as coffee, take up the largest percentage of traceable plant families collected Indonesia and Papua New Guinea. In the early 1880s, the Dutch East Indies was one of the major producers of coffee in the world (Fernando, 2003). Since most of the coffee was exported to the Netherlands, detailed knowledge about coffee plants was essential for colonial administrators and botanists in Europe and insular Southeast Asia (Schoor, 2012).

Orchids also represent an important section, comprising 12,65% of traded species in our dataset. Furthermore, other plants seen in the interface such as *Moraceae* (most commonly known as figs), *Lauraceae* (laurel), *Leguminosae* (legumes and peas), *Myr-*

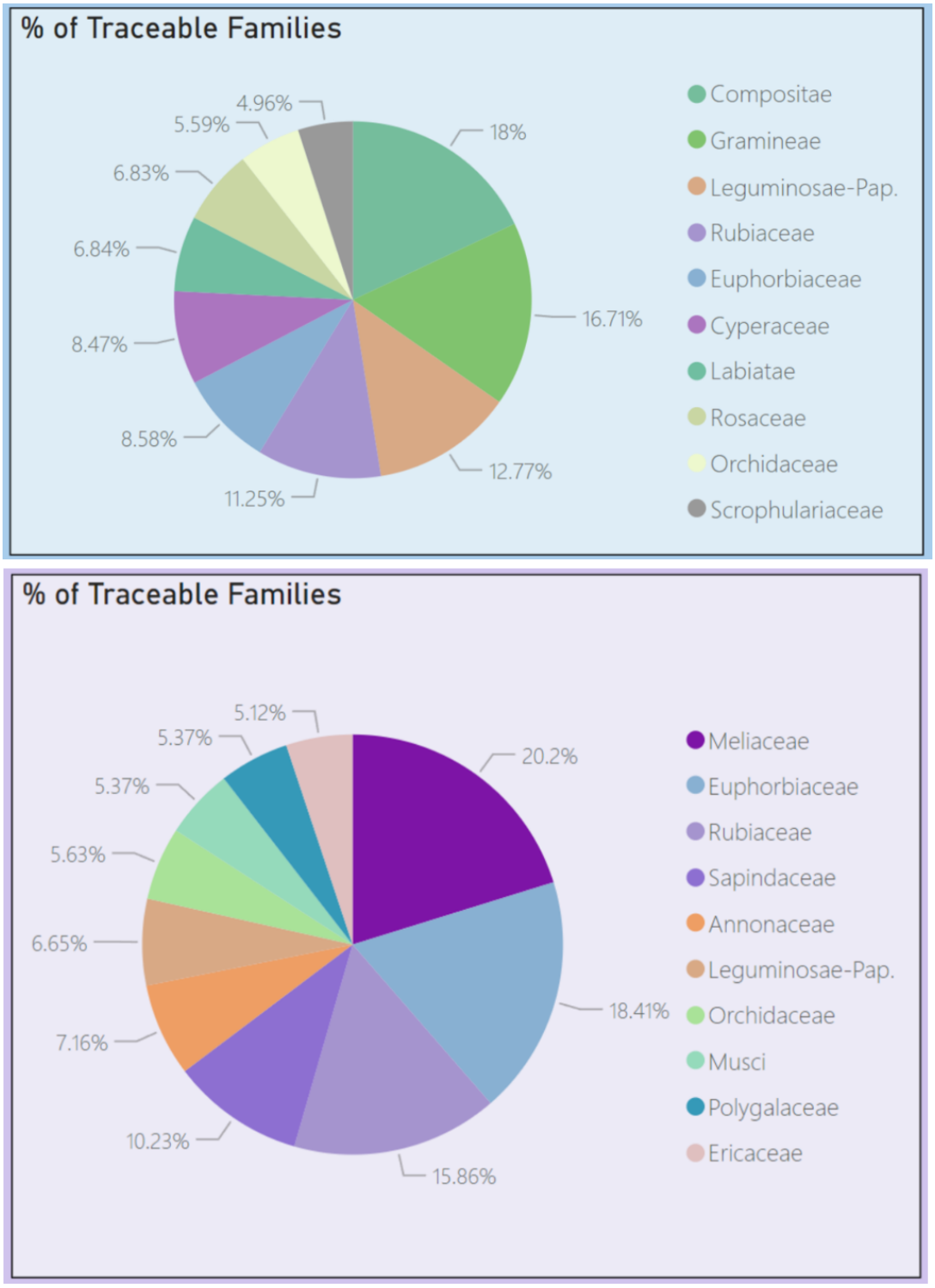

Figure 4: Percentage of traceable families from all the countries (top) and from Indonesia and Papua New Guinea (bottom).

taceae (clove) and *Meliaceae* (cotton fruits, commonly consumed in South East Asian regions) were collected during colonial times. According to our data, the Naturalis collection entails ca. 20.000 Moracea specimens from Indonesia with a collection peak in 1961. Legumes, peas and other edible plants were also accumulated, particularly between the nineteenth and twentieth centuries, matching the timeline showcased on the interface (Wasino et al., 2017, pp. 265-267). Observing the bottom of Figure 4, *Meliaceae*, most commonly known as neem, was widely traded and forms an important part of the museum's botanical collection. Its consumption took many forms, particularly for "pressed seed oil, mainly used as insecticide, but also for cosmetic, medicinal and agricultural uses" (Sujarwo et al., 2016, p.186). Data with respect to other species, such as *Labiatae* (mint), and *Compositae* (daisies) also show interesting spikes which would require further contextualization. Taken together, the in-depth examination of data related to specific plant families illustrate how the collection's composition was also shaped by economic motives geared towards extracting agricultural resources from areas which fell under Dutch colonial rule.

6 Discussion and Conclusion

This project thus contributes to filling a gap in historical research regarding the colonial legacy of Dutch botanical collections, highlighting the intertwined nature of scientific endeavors and history. Furthermore, it addresses a gap in visualization-oriented digital humanities research by providing a concrete example of how easy-to-use computational techniques can help in understanding the cultural histories of scientific collections in an interactive way, and can thereby help decolonization of natural history collections. Specifically, we have demonstrated how simple and interactive visualizations can help researchers to develop a better understanding of a collection's colonial bias and how this bias has evolved historically (Nelson and Ellis, 2019). Our paper presents an opportunity for natural history museums to make their data more accessible through a user-friendly and visual tool that is easy to deploy and does not consume much computational resources.

Other Dutch museums such as the Wereldmuseum have addressed colonial Dutch history by decolonizing their collections and exhibits.[5] Their work has shown how addressing colonial history can lead to making collections and exhibitions more inclusive and reflective of diverse cultural histories. Enabling museum visitors and users of scientific collection-related datasets to study colonial provenance acknowledges the colonial biases of natural history collections. Tools such as Power BI play a crucial role in reconnecting datasets to their cultural and historical roots by enabling dynamic visualizations that reveal patterns, biases, and historical trajectories otherwise hidden in raw data. By integrating spatial, temporal, and categorical filters, these tools allow scientists to trace the provenance of specimens, historians to contextualize collection practices within colonial frameworks, and a broader multidisciplinary audience to engage with data through an intuitive interface. The ability to filter by collector, location, time period, inter alia, highlights how scientific knowledge production was shaped by historical power dynamics, making the tool particularly useful, i.e. for museum professionals and researchers investigating provenance. At the same time, its accessibility, which does not require specialized coding skills, expands its reach to the general public, fostering greater engagement with the ethical dimensions of natural history collections. By bridging these disciplinary and public divides, Power BI

[5] See also here: ../why-exhibition-about-our-colonial-inheritance, accessed 10 October 2024.

serves as a vital instrument for uncovering, analyzing, and sharing the layered histories embedded within scientific datasets. Thus, our interactive tool links the Netherlands' colonial past with Naturalis' scientific plant collection through multi-layered visualizations. Additionally, it helps researchers to uncover important stories of how the collections have been brought together, contributing to a more nuanced interpretation of their collections' evolution over a longer period. The initial colonial context behind botanical acquisitions set the trajectory for future Dutch botanical endeavors in the various countries, even after they gained independence.

Our dashboards - and our focus on the two former Dutch colonies - allows users to employ a historical lens to reveal dynamic relationships between natural history collection patterns and broader socio-political contexts. Since the interactive visualizations allow natural history collections to be studied in terms of their movement (via the map tool) over time (via the timeline) and allow for detailed analysis (via the metadata filters), it can be used to investigate a range of questions related to the cultural histories of scientific collections. The flexibility it allows the user, its ability to make data more accessible, its potential use in flagging missing data and the clear ability to reveal information that would otherwise be hidden shows that this interactive visualization can best contribute to understanding the cultural histories of scientific collections in a way that static visualizations otherwise would not. As depicted, the visualization tool enhances the collections' accessibility through visual reading, with the potential to reveal insights otherwise hidden and uncover data-related concerns. The tool thus aids biologists and those interested in tracing the provenance, geographical spread of natural objects in collections and the ethics of obtaining these specimens in the context of colonial power asymmetries.

However, certain limitations remain. Firstly, this study focuses solely on the botany subsection of Naturalis, leaving the broader applicability of our approach to the entire natural history collection untested, particularly in terms of merging multiple datasets and assessing Power BI's scalability. Due to the scope of this research, we were unable to evaluate the generalization of our experiment by incorporating additional large datasets, an avenue we intend to explore in future work. Expanding this approach to include datasets from different institutions would provide a more thorough assessment of Power BI's ability to handle larger and more diverse data sources. While the platform is well-suited for structured data visualization, its performance may be challenged by increased data volume and variability, potentially necessitating extensive preprocessing or alternative tools for optimal scalability. Future research should examine how integrating datasets from various natural history museums, each with distinct formats and metadata standards, influences the visualization process while ensuring that insights into colonial biases remain clear and accessible. Secondly, since the Naturalis botany dataset originates from the GBIF platform, adjustments may be required to accommodate datasets from other institutions, as formatting conventions are unlikely to align seamlessly. Lastly, leveraging Power BI for such visualizations may demand significant data preprocessing, which could present a barrier for users with limited data management skills.

Nonetheless, this visualization tool provides a starting point for tracing the histories of any museum collection, provided they have the data. The *recordedBy* metadata, for instance - through which explorations and records made by botanists can be tracked and mapped - is an interesting starting point to understand the specific acquisition histories of certain specimens. For example, Carl Ludwig Blume - whose records show that 95,77% of the specimens he recorded belong to Indonesia (see Figure 5)

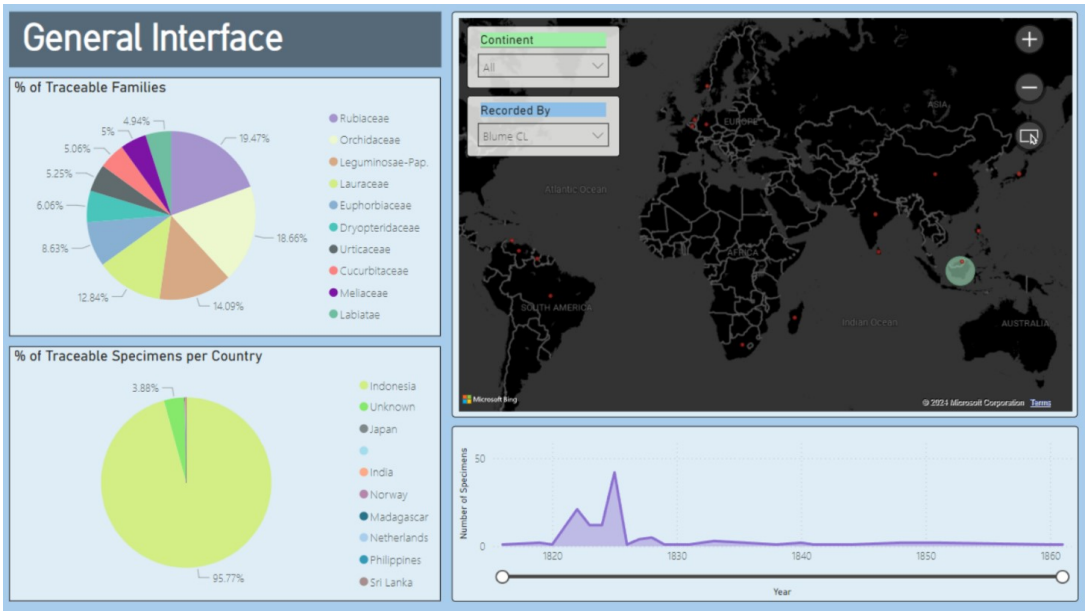

Figure 5: General Interface with the Recorded By filter set to Blume CL, showing that the majority of specimens he recorded came from Indonesia.

- is a relevant finding, indicating the Dutch botanical connection to their colonies. Blume was the founder of 's Rijksherbarium (National Herbarium), an institutional predecessor of what is now museum Naturalis. The sheer amount of orchids taken from Indonesia by a single researcher is also an interesting finding in itself, opening a Pandora's box of other insightful questions. Thus, the scope of such a tool can undoubtedly be extended to address other questions of contemporary relevance. While in the nineteenth century, orchids were primarily seen as scientific study objects and luxury products from the East and West Indies, they are now a globally traded horticultural commodity whose presence in almost every European household has severe environmental consequences (Soode et al., 2015).

7 Data Access

The Power BI file used for this paper can be downloaded through Zenodo: `https://doi.org/10.5281/zenodo.13958314`.

References

Godwin Ubong Akpan, Isah Mohammed Bello, Kebba Touray, Reuben Ngofa, Daniel Rasheed Oyaole, Sylvester Maleghemi, Marie Babona, Chanda Chikwanda, Alain Poy, Franck Mboussou, et al. Leveraging polio geographic information system platforms in the african region for mitigating covid-19 contact tracing and surveillance challenges. *JMIR mHealth and uHealth*, 10(3):e22544, 2022.

Mireia Alcantara-Rodriguez, Tinde Van Andel, and Mariana Françozo. Nature portrayed in images in Dutch Brazil: Tracing the sources of the plant woodcuts in the Historia Naturalis Brasiliae (1648). *bioRxiv*, pages 2022–10, 2022.

Tinde van Andel. *Open the treasure room and decolonize the museum*. Universiteit Leiden, 2017.

Jack Ashby and Rebecca Machin. Legacies of colonial violence in natural history collections. *Journal of Natural Science Collections*, 8:44–54, 2021.

Zaheer Baber. The plants of empire: Botanic gardens, colonial power and botanical knowledge. *Journal of Contemporary Asia*, 46(4):659–679, 2016.

Vijay Barve and Javier Otegui. bdvis: visualizing biodiversity data in r. *Bioinformatics*, 32(19):3049–3050, 2016.

Tiziana N Beltrame, Elena Canadelli, and Luca Tonetti. The natures of digital practices: People, objects, and data mobilities in natural history collections. *Nuncius*, 39(3): 741–758, 2024.

Elizabeth H Boakes, Philip JK McGowan, Richard A Fuller, Ding Chang-qing, Natalie E Clark, Kim O'Connor, and Georgina M Mace. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS biology*, 8(6):e1000385, 2010.

Peter Boomgaard. The making and unmaking of tropical science: Dutch research on indonesia, 1600-2000. *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia*, 162(2):191–217, 2008.

Marieke Borren and Caroline Drieënhuizen. Dossier: The coloniality of natural history collections. *Locus: Tijdschrift voor Cultuurwetenschappen*, 2022.

Lissa Breugelmans and Maarten Trekels. Implementation experience report for the developing latimer core standard: The dissco flanders use-case. *Biodiversity Information Science and Standards*, 7:e113766, 2023.

Falko T Buschke, Claudia Capitani, El Hadji Sow, Yvonne Khaemba, Beth A Kaplin, Andrew Skowno, David Chiawo, Tim Hirsch, Elizabeth R Ellwood, Hayley Clements, et al. Make global biodiversity information useful to national decision-makers. *Nature Ecology & Evolution*, 7(12):1953–1956, 2023.

Anne Coote, Alison Haynes, Jude Philp, and Simon Ville. When commerce, science, and leisure collaborated: the nineteenth-century global trade boom in natural history collections. *Journal of Global History*, 12(3):319–339, 2017.

Jeroen Creuwels. Naturalis Biodiversity Center (nl) - Botany Leiden. Occurrence dataset. *Naturalis Biodiversity Center*, 2015.

Helen Anne Curry, Nicholas Jardine, James Andrew Secord, and Emma C Spary. *Worlds of natural history*. Cambridge University Press, 2018.

Subhadra Das and Miranda Lowe. Nature read in black and white: decolonial approaches to interpreting natural history collections. *Journal of Natural Science Collections*, 6(4):4–14, 2018.

Verónica Díez Díaz, Sara Akhlaq, Katja Kaiser, Ina Heumann, and Daniela Schwarz. Digitization as a research methodology in colonial natural history collections. *Nature Reviews Biodiversity*, pages 1–2, 2025.

DiSSCo. Nationaal natuurhistorisch collectieoverzicht. *https://app.powerbi.com/view?r=eyJrIjoiNDQ1YmQzMzMtNmY5YS00MDQzLWI5M2YtNmRhOTM2MTg2NTU0IiwidCI6IjhjZDI0OTg0LTBhYTMtNGZjNS1iMD liLTRkNmVjZmFhNThmYiIsImMiOjl9, accessed 19 October 2024*, 2023.

Richard Drayton. Nature's government: Science, imperial britain, and the 'improvement'of the world, 2000.

Joshua A Drew, Corrie S Moreau, and Melanie LJ Stiassny. Digitization of museum collections holds the potential to enhance researcher diversity. *Nature ecology & evolution*, 1(12):1789–1790, 2017.

Caroline Drieënhuizen and Fenneke Sysling. Java man and the politics of natural history: An object biography. *Bijdragen tot de taal-, land-en volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia*, 177(2-3):290–311, 2021.

Felix Driver, Mark Nesbitt, and Caroline Cornish. *Mobile museums: Collections in circulation*. UCL Press, 2021.

Déborah Dubald and Catarina Madruga. Introduction: Situated nature: Field collecting and local knowledge in the nineteenth century. *Journal for the History of Knowledge*, 3 (1):1–11, 2022.

Jim Endersby. *Orchid: A cultural history*. University of Chicago Press, 2016.

Xiao Feng, Brian J Enquist, Daniel S Park, Brad Boyle, David D Breshears, Rachael V Gallagher, Aaron Lien, Erica A Newman, Joseph R Burger, Brian S Maitner, et al. A review of the heterogeneous landscape of biodiversity databases: Opportunities and challenges for a synthesized biodiversity knowledge base. *Global Ecology and Biogeography*, 31(7):1242–1260, 2022.

M.R. Fernando. *Coffee Cultivation in Java, 1830–1917*, pages 157–172. 06 2003. ISBN 9780521818513. doi: 10.1017/CBO9780511512193.017.

Nico M Franz and Beckett W Sterner. To increase trust, change the social design behind aggregated biodiversity data. *Database*, 2018:bax100, 2018.

Emilio García-Roselló, Jacinto González-Dacosta, and Jorge M Lobo. The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. *Biological Conservation*, 283: 110118, 2023.

Andrew Goss. Decent colonialism? Pure science and colonial ideology in the Netherlands East Indies, 1910–1929. *Journal of Southeast Asian Studies*, 40(1):187–214, 2009.

Andrew Goss. *The floracrats: State-sponsored science and the failure of the enlightenment in Indonesia*. Univ of Wisconsin Press, 2011.

Andrew Goss. Reinventing the kebun raya in the new republic: Scientific research at the Bogor Botanical Gardens in the age of decolonization. *Studium*, 11(3), 2018.

Andrew Goss. Decolonizing botany: Indonesia, unesco, and the making of a global science. *Journal of the History of Biology*, 56(3):495–523, 2023.

Carl Haarnack. Met andere ogen: Over de slavernijtentoonstelling in het Rijksmuseum. *Jaarboek De Achttiende Eeuw*, 53(1):193–199, 2021.

Brandon P Hedrick, J Mason Heberling, Emily K Meineke, Kathryn G Turner, Christopher J Grassa, Daniel S Park, Jonathan Kennedy, Julia A Clarke, Joseph A Cook, David C Blackburn, et al. Digitization and the future of natural history collections. *BioScience*, 70(3):243–251, 2020.

P. Hiepko. The collections of the Botanical Museum Berlin-Dahlem (b) and their history. *Englera*, 7:219–252, 1987.

Roos Hopman. Snails, time, data: On the politics of mass-digitization and the possibility of data drift. *Big Data & Society*, 11(3):20539517241267760, 2024.

Alice C Hughes, James B Dorey, Silas Bossert, Huijie Qiao, and Michael C Orr. Big data, big problems? how to circumvent problems in biodiversity mapping and ensure meaningful results. *Ecography*, page e07115, 2024.

Dominik Hünniger. Visible labour? Productive forces and imaginaries of participation in european insect studies, ca. 1680–1810. *Berichte zur Wissenschaftsgeschichte*, 44(2): 180–210, 2021.

Sharif Islam, Andreas Weber, and Erzsébet Tóth-Czifra. From green deal to cultural heritage: Fair digital objects and european common data spaces. *Research Ideas and Outcomes*, 8:e93815, 2022.

Hans J Jacobs and André Koch. On the discovery and scientific description of the Emerald Tree Monitor, Varanus prasinus (Schlegel, 1839). *Bibliotheca herpetologica*, 15(7):61–76, 2021.

Hans J Jacobs and Glenn M Shea. Eight weeks in lobo bay. the natuurkundige commissie on new guinea in 1828. i. scincus and centroplites (scincidae). *Bibliotheca herpetologica*, 16(6):48–81, 2022.

Kirk R Johnson, Ian FP Owens, and Global Collection Group. A global approach for natural history museum collections. *Science*, 379(6638):1192–1194, 2023.

Katja Kaiser, Ina Heumann, T Nadim, H Keysar, M Petersen, M Korun, and F Berger. Promises of mass digitisation and the colonial realities of natural history collections. *Journal of Natural Science Collections*, 11:13–25, 2023.

Kelsey Leonard. Decolonizing botanical gardens. *Qualitative Research Journal*, 24(5): 536–554, 2024.

Carla Maldonado, Carlos I Molina, Alexander Zizka, Claes Persson, Charlotte M Taylor, Joaquina Albán, Eder Chilquillo, Nina Rønsted, and Alexandre Antonelli. Estimating species diversity and distribution in the era of b ig d ata: to what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8):973–984, 2015.

Maarten Manse. Kennis is macht: de veelzijdige expedities van botanicus Pieter Willem Korthals (1807–1892). *Studium*, 6(1), 2013.

Asla Medeiros e Sá, Franklin Alves de Oliveira, Bruno Schneider, Karina Rodriguez Echavarria, and Cristiana Silveira Serejo. Visually overviewing biodiversity open data digital collections. In *Proceedings of the Symposium on Open Data and Knowledge for a Post-Pandemic Era ODAK22, UK*, pages 1–8. BCS Learning & Development, 2022.

Yamikani Mgusha, Deliwe Bernadette Nkhoma, Msandeni Chiume, Beatrice Gundo, Rodwell Gundo, Farah Shair, Tim Hull-Bailey, Monica Lakhanpaul, Fabianna Lorencatto, Michelle Heys, et al. Admissions to a low-resource neonatal unit in malawi using a mobile app and dashboard: a 1-year digital perinatal outcome audit. *Frontiers in Digital Health*, 3:761128, 2021.

Eulàlia Gassó Miracle. *Coenraad Jacob Temminck and the emergence of systematics (1800–1850)*. Brill, 2021.

Ryan S Mohammed, Grace Turner, Kelly Fowler, Michael Pateman, Maria A Nieves-Colón, Lanya Fanovich, Siobhan B Cooke, Liliana M Dávalos, Scott M Fitzpatrick, Christina M Giovas, et al. Colonial legacies influence biodiversity lessons: how past trade routes and power dynamics shape present-day scientific research and professional opportunities for caribbean scientists. *The American Naturalist*, 200(1): 140–155, 2022.

Johannes Müller. Naming the world: Pieter Bleeker's travels and the challenges of archipelagic biodiversity. *Animals in Dutch travel writing, ed. by R.A.M. Honings and E. Beek*, pages 79–98, 2023.

Tahani Nadim, Mareike Vennen, Ina Heumann, and Filippo Bertoni. Logistical natures: Trade, traffics, and transformations in natural history collecting. *Historical Studies in the Natural Sciences*, 54(2):125–134, 2024.

Gil Nelson and Shari Ellis. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B*, 374(1763):20170391, 2019.

Daniel S Park, Xiao Feng, Shinobu Akiyama, Marlina Ardiyani, Neida Avendaño, Zoltan Barina, Blandine Bärtschi, Manuel Belgrano, Julio Betancur, Roxali Bijmoer, et al. The colonial legacy of herbaria. *Nature Human Behaviour*, 7(7):1059–1068, 2023.

Victoria Pickering. Mobilising historical botanical data as research. *Nuncius*, 39(3): 759–774, 2024.

Aníbal Quijano. Coloniality and modernity/rationality. *Cultural studies*, 21(2-3): 168–178, 2007.

Willem Jan van der Schoor. Zuivere en toegepaste wetenschap in de tropen: biologisch onderzoek aan particuliere proefstations in Nederlands-Indië 1870-1940. *Ph. D. Thesis*, 2012.

Eveli Soode, Paul Lampert, Gabriele Weber-Blaschke, and Klaus Richter. Carbon footprints of the horticultural products strawberries, asparagus, roses and orchids in germany. *Journal of Cleaner Production*, 87:168–179, 2015.

Lise Stork, Andreas Weber, Eulàlia Gassó Miracle, Fons Verbeek, Aske Plaat, Jaap van den Herik, and Katherine Wolstencroft. Semantic annotation of natural history collections. *Journal of Web Semantics*, 59:100462, 2019.

Lise Stork, Andreas Weber, Jaap van den Herik, Aske Plaat, Fons Verbeek, and Katherine Wolstencroft. Large-scale zero-shot learning in the wild: Classifying zoological illustrations. *Ecological informatics*, 62:101222, 2021.

Wawan Sujarwo, Ary P Keim, Giulia Caneva, Chiara Toniolo, and Marcello Nicoletti. Ethnobotanical uses of neem (azadirachta indica a. juss.; meliaceae) leaves in Bali (Indonesia) and the Indian subcontinent in relation with historical background and phytochemical properties. *Journal of Ethnopharmacology*, 189:186–193, 2016.

F. Sysling, H. Vellinga, S. Kompier, and T. Niederer. Who did all the work. *https://whodidallthework.nl, accessed 19 October 2024*, 2024.

Marieke Sophia van de Loosdrecht, Nicholaas M Pinas, Jerry R Tjoe Awie, Frank FM Becker, Harro Maat, Robin van Velzen, Tinde van Andel, and M Eric Schranz. Maroon rice genomic diversity reflects 350 years of colonial history. *bioRxiv*, pages 2024–05, 2024.

Lorella Viola. Networks of migrants' narratives: a post-authentic approach to heritage visualisation. *ACM Journal on Computing and Cultural Heritage*, 16(1):1–21, 2023.

M Wasino, Hartatik Endah Sri, et al. Peasant economy under the plantation capitalism (the study of peasant agricultural plantation in java at the nineteenth to the beginning of twentieth century). *Man in India International Journal of Anthropology. Vol. 97 No. 5, 2017.*, 2017.

Andreas Weber. A garden as a niche: Botany and imperial politics in the early nineteenth century Dutch empire. *Studium*, 11(3):178–190, 2018.

Andreas Weber and Esther Turnhout. A langur from sumatra: Digital futures, material presents and colonial pasts. *Nuncius*, 39(3):775–788, 2024.

Andreas Weber and Robert-Jan Wille. Laborious transformations: Plants and politics at the bogor botanical gardens. *Studium*, 11(3):169–177, 2018.

Robert-Jan Wille. *Mannen van de microscoop: De laboratoriumbiologie op veldtocht in Nederland en Indië, 1840-1910*. Vantilt, 2019.

Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE transactions on visualization and computer graphics*, 25(6):2311–2330, 2018.

Pieter van Wingerden. Science on the edge of empire: E.A. Forsten (1811–1843) and the Natural History Committee (1820–1850) in the Netherlands Indies. *Centaurus*, 62(4):797–821, 2020.

Pieter van Wingerden. *The Natuurkundige Commissie in the Netherlands Indies (1820-1850)*. Brill, 2025.