Biological Journal of the Linnean Society, 2026, **147**(**1**), blaf115
https://doi.org/10.1093/biolinnean/blaf115

OXFORD

Original Article

Materiality and context of specimen labels: exposing frictions between local and global priorities in the preservation of 'non-prototypical' records

Thiago S. R. Silva[1,2,*], Ana Carolina A. Neundorf[1], Rodrigo M. Feitosa[1]

[1]Departamento de Zoologia, Universidade Federal do Paraná, C.P. 19020, CEP 81531-980, Curitiba, PR, Brazil
[2]Zoology Unit, Finnish Museum of Natural History, Pohjoinen Rautatiekatu 13, 00100, Helsinki, Uusimaa, Finland

[*]Corresponding author. Zoology Unit, Finnish Museum of Natural History, Pohjoinen Rautatiekatu 13, 00100, Helsinki, Uusimaa, Finland.
E-mail: tsranzanidasilva@gmail.com

ABSTRACT

Specimen labels are frequently considered metadata by biologists worldwide, acquiring the form of a source of 'primary biodiversity data'. Frequently, the materiality of specimen labels is considered problematic, given the difficulties in translating the data contained on them to machine-readable formats, possibly hampering the comparability of biological data. However, this framework frequently ignores the fact that the materiality of specimen labels is a manifestation of the historical and social frictions of the agents involved in documenting biodiversity. Here, we show how the materiality of specimen labels reflects the epistemic frictions of agents responsible for fieldwork and laboratory documentation. Our results show how the demands of recorders and collectors are frequently manifested physically during specimen label creation in different moments of its life-cycle, while exposing the socio-cultural struggles in preparing and preserving this type of record, and the institutional constraints in their curation, affecting policy-making at the institutional level.

Keywords: biological collections; archival research; ethnography; material records; curation policies

INTRODUCTION

Specimen labels are frequently considered metadata by biologists worldwide (Johnson *et al.* 2023), serving as sources of 'primary biodiversity data', a general term for information documenting the biodiversity of the planet, where each record represents the presence of a particular organism at a given location and time (Holetschek *et al.* 2012). Because of their relevance to fields such as macroecology and macroevolution, the materiality of specimen labels and their inherited idiosyncrasies are often considered problematic, particularly when translating label content into machine-readable formats (Vollmar *et al.* 2010, Thessen *et al.* 2012). Some researchers further classify specimen labels as secondary physical objects (Hardisty *et al.* 2022). Within such frameworks, specimen labels are frequently reduced to 'data chunks' [as understood in the RFC 4960 SCTP standard (Stewart 2007)], with museums and collections viewed as vaults containing 'minable' data (Graham *et al.* 2004, Rudolph and Wiman 2023).

This data-centric perspective, however, obscures the material and social dimensions of label production, use, and preservation. Additionally, the apparent neutrality of 'data chunks' conceals the selection, translation, and negotiation embedded in their creation (Gitelman and Jackson 2013). As argued by Bowker (2000) and Bates *et al.* (2016), data are not simply extracted but are actively shaped by institutional, historical, and technological contexts. The concept of the 'data journey' highlights how data gain meaning as they move through acquisition, processing, databasing, and dissemination, rather than functioning as atomized carriers of information.

From an archival perspective, specimen labels can be understood through Geoffrey Yeo's paradigm (Yeo 2007, 2008), defining them as records by three key features: (i) their persistence beyond the moment of creation; (ii) their nature as representations of activities; and (iii) their creation by participants or observers with first-hand knowledge. Thus, these labels are not mere detached informational units but persistent, first-hand accounts of scientific work. Yet, persistence is contingent rather than intrinsic, because specimen labels are subject to material degradation, and their continuity depends on institutional practices of preservation and duplication. This emphasizes that persistence is bound up with care and preservation, not durability alone. Within Yeo's typology,

Received 01 May 2025; revised 09 September 2025; accepted 26 September 2025

specimen labels occupy a liminal position as 'non-prototypical' records, situated between formal scientific documentation and personal and work-based artefacts.

Through this perspective, the procedures involved in preparing specimen labels can be aligned with the concept of data journeys. Labels are material products shaped by socio-technical systems rather than solely instruments of data management. Their physical form reflects labour-intensive processes of observation, negotiation, and inscription, conditioned by disciplinary standards, institutional protocols, and the constraints of field and laboratory work, and collection management. Treating them as epistemically laden material artefacts highlights how the production of knowledge in biology is structured by context-specific norms and infrastructures.

The second conceptual anchor of this study draws from commons scholarship. Biodiversity and its associated data are increasingly framed as global commons: shared resources with transnational significance, whose conservation and use involve a multiplicity of actors with uneven stakes, responsibilities, and access (Horning 2010, Berkes 2006). This framework helps to illuminate the friction between two conflicting pressures: (i) the global imperative to make biodiversity data open, interoperable, and FAIR (Findable, Accessible, Interoperable, and Reusable); and (ii) local constraints and priorities that shape how records are created, maintained, and valued within regional or national institutions. Rather than tracing a linear flow from local to global, in this study we examine the frictions that emerge at their intersection, particularly concerning funding policies, documentation protocols, and institutional roles in record-making.

In museum studies, material entities (understood in this field as material culture) are seen as a system of interdependent artefacts, where the value of each artefact results solely from the simultaneous presence of other artefacts (Pearce 1989). These material entities are often manifested as tangible outcomes of human demands, shaped by specific human forces, selected from social and 'natural environments', and subsequently preserved, decoded, and used for specific purposes (Schreiner 1988, van Mensch 1990). According to Ilerbaig (2010), this two-stage view of material entities can be parallelized to the life-cycle model of archival records, where these objects live a 'double-life' of sorts, initially in connection with human activity and then as documentation of such activity. This perspective, although fundamentally human-centred, because it focuses on the activities and perspectives of the human agents involved in the collection event rather than on representing the environment with 'accuracy', is also a way of obscuring the context from which the record was created.

The handling of archival records (government papers, diaries, newspapers, etc.) is widely seen as the hallmark of historians and archivists. At the same time, biologists are often identified with sampling techniques, specimen examination, and aggregation of atomized data. However, biologists routinely use archival records not only for writing articles and to analyse explanatory potentialities of biological phenomena, but also to understand what Karen Barad (2007) describes as ethico-onto-epistemological entanglement, i.e. the inseparability of ethics (how we ought to act), ontology (what we believe exists), and epistemology (how we know) in the practices of science. From this perspective, the material entities, the generation of meaning, and the social dynamics of a given collective are deeply intertwined, exposing that scientific practices are intrinsically ethical and political. Recent work has approached specimen labels as archival records beyond their role as data providers, drawing attention to their historiographic, ethnographic, and conservation significance (Corrado *et al.* 2015, Wood 2025).

Building on this framework, in the present study we examine specimen labels not merely as instruments of data management but as material records embedded in socio-technical and institutional practices. Drawing on ethnographic research conducted in an entomology laboratory, the analysis addresses how labels are designed, produced, and interpreted, with particular attention to the epistemic frictions that arise when biological observations are translated into textual form. These frictions encompass decisions about language, format, taxonomic conventions, spatial constraints, and concerns over long-term legibility.

By emphasizing the material function of specimen labels, this study contributes to scholarship interrogating the socio-material constitution of scientific artefacts. Instead of treating labels as neutral carriers of information, it emphasizes their role as epistemic artefacts shaped by institutional constraints, disciplinary norms, and labour practices. This perspective also exposes the ethical and political dimensions of record-making, aligning with Barad's (2007) notion of ethico-onto-epistemological entanglement.

Recognizing specimen labels as archival records and products of specific cultural, institutional, and scientific practices enables a more nuanced understanding of their role in biology. Reframing them from passive data containers to historically situated and materially grounded records makes it possible to account for the labour, social dynamics, and communicative struggles involved in their production and conservation and to inform more ethically attuned and context-sensitive policies for managing biological collections.

Given the explanation above, the aim of this study was to reframe perspectives on specimen labels in biological collections, shifting their position from 'data chunks' to a position as material records of activities, based on an ethnographic study of an entomology laboratory, specifically analysing the production of specimen labels. We discuss how the materiality of specimen labels reflects the epistemic frictions of agents responsible for fieldwork and laboratory documentation, the socio-cultural struggles in preparing and preserving this type of record, and the institutional constraints in their curation. Finally, we emphasize that specimen labels, in their present form, are not merely a matter of necessity but a natural physical manifestation of the day-to-day dynamics within biological collections. As such, they should be accounted for in policies aimed at conserving these collections.

MATERIALS AND METHODS

Fieldwork and laboratory research were conducted between 2017 and 2024 in Brazil and Hong Kong using a multi-sited ethnographic approach. Data collection methods included participant observation, unstructured note-taking, and document analysis in both field and institutional settings.

Two ant collection events took place in the Parque Estadual das Lauráceas, Paraná, Brazil, in May 2017 and July 2019. During

these events, one investigator (R.M.F.) recorded collection parameters including geolocation, sampling method, date, and participating agents. This information was documented in a personal field notebook and, when relevant, informed the creation of provisional specimen labels used on site. A second investigator (T.S.R.S.) documented team interactions, including deliberations about site selection, decisions on sampling techniques, and practical challenges, such as label legibility. These observations were recorded as unstructured notes, which also included reflections on the epistemic implications of different labeling practices.

Laboratory work in Brazil was conducted intermittently from 2019 to 2024 at the Laboratório de Sistemática e Biologia de Formigas (LSBF), Universidade Federal do Paraná Curitiba, Brazil. Throughout this period, label production adhered to protocols established within the LSBF. One investigator (A.C.A.N.) generated permanent labels from parameters obtained during field campaigns and recorded notes on the procedures involved in their preparation.

Additional ethnographic research was carried out at the Insect Biodiversity and Biogeography Laboratory (IBBL) and the Hong Kong Biodiversity Museum (HKBM), both at the University of Hong Kong. Observations were conducted in 2022 and 2023. During this period, one investigator (T.S.R.S.) prepared specimen labels using personal protocols and maintained a work diary documenting institutional practices related to specimen labelling, digitization, and collection curation. Observations focused on the availability and sourcing of label materials, institutional preferences for label types (e.g. minimalist code labels vs. traditional specimen labels), and formatting conventions. Attention was also given to institutional procedures governing the acquisition of materials and reimbursement processes, with specific emphasis on how these practices affected access to archival-quality supplies and other resources required for label production. In both laboratory contexts, instances of material deterioration (such as discoloration, tearing, and partial loss of labels) were also noted, highlighting the vulnerability of records to processes of decay alongside their preservation.

Notes were analysed using a reflexive ethnographic approach, emphasizing the embedded nature of the researchers within the environments studied. Field notes, label designs, and institutional documents were examined to identify epistemic assumptions, frictions, and material constraints shaping the creation and standardization of specimen labels. These empirical observations were interpreted in dialogue with conceptual perspectives from archival studies and commons scholarship to examine how labelling practices reflect and mediate broader scientific, institutional, and political commitments in biodiversity research.

RESULTS

Representations of labels through time: differential procedures of record-making in ant collections

Specimen labels vary in format and production methods throughout the record preparation process, influenced by the storage medium, the purpose of their creation, the materials available, and the taxonomic group being sampled.

To illustrate these points, we will transcribe two collection events conducted by members of the LSBF at different times. These events took place in 2017 and 2019 at the Parque Estadual das Lauráceas, whose headquarters are in the municipality of Adrianópolis (Fig. 1).

Researchers tend to prepare at least two different types of specimen labels, depending on the goal of the collection trip. Given that LSBF conducts taxonomy and ecology studies, the process of creating specimen labels varies in the field. In taxonomy-focused studies, researchers document the sampling technique used and the specific location where the specimens were collected (e.g. leaf litter extraction on roots of '*canela-fogo*'). In ecological studies, researchers often use codes to link specimens collected from the same sampling unit (e.g. S2T5W10 for site 2, transect 5, Winkler sample 10).

In both cases, labels are generally prepared in the field using parchment paper as the medium and a pencil or 0.05–0.1 mm pin pen for recording information. These provisional or field labels are prepared for individual vials and are characterized by relatively informal documentation practices: low-quality papers and pencils might be used, and linguistic conventions tend to be flexible, with agents often identified by given names or nicknames (e.g. 'Rodrigo', 'Carol', or 'Thi'). This contrasts with the 'documentation responsibility' associated with permanent labels, where linguistic variation gives way to more or less standardized forms. The dissociation between these two modes of inscription (provisional and permanent) is thus materialized in both the physical properties of the labels and the denominative practices they embody.

After the collection, the specimens are matched with their respective labels and placed in the vials. The full details of the sampling event are recorded on an A4 sheet of paper and transcribed using a 0.1 mm pin pen. To facilitate ease of writing and reading, the locality and other relevant information are often abbreviated (Fig. 2; e.g. P.E. das Lauráceas).

At the end of the field trip, the samples (along with their respective records) are taken to the laboratory for processing. In the laboratory, the provisional labels are replaced by permanent labels, which will accompany the specimens when incorporated into the collection. The preparation of the permanent label is important and deserves attention and care, given that this information gives 'meaning' to the specimen, because a specimen without provenance data, i.e. without associated information, does not have 'scientific value'. Upon arrival at the laboratory, each sample is processed, and the specimens obtained are grouped into subsamples, whose categorization is defined through the criterion of morphological similarity. This procedure is carried out on a bench with space for a stereomicroscope, glass vials for sample preservation, conservation liquid, and the samples obtained in the field (Fig. 3). This task is normally conducted by a laboratory member with the required training and knowledge to execute the task (in our case, A.C.A.N.).

The permanent labels are prepared in a text editor according to the preference of the preparer and have approximate dimensions of 0.80 cm in height and 2 cm in width. The preferred font should be Arial or Arial Narrow (some researchers use Helvetica), size 4 or 4.5. The font size can vary according to the number of lines and information on the labels, with five lines being the maximum number to be used. The information inserted in the permanent label should be, respectively: country (bold and upper case letters, sometimes written as an abbreviation, such as USA), followed by

Figure 1. Expedition to Parque Estadual das Lauráceas in 2017 and 2019. A, overview of the park from the headquarters. B, camp base, where sample preprocessing and preparation of provisional specimen labels took place. C, equipment and infrastructure used for sample preprocessing and provisional specimen label preparation. D, LSBF members preprocessing samples and preparing provisional specimen labels during the 2017 expedition. E, LSBF members preprocessing samples and preparing provisional specimen labels during the 2019 expedition. Images by Rodrigo Feitosa. Abbreviation: LSBF, Laboratório de Sistemática e Biologia de Formigas.

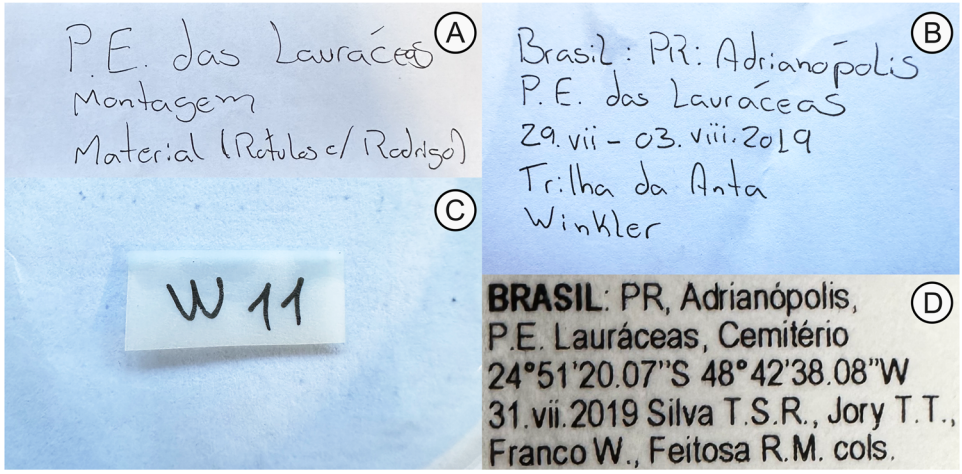

Figure 2. Models of labels used in different moments. A, preprocessing label. B, provisional label. C, additional information associated with the provisional label containing the sampling point and technique used. D, permanent specimen label. Images by Thiago Silva.

(colon delimited) the province/department or equivalent (which can be an abbreviation, if broadly known and used, e.g. PR standing for the state of Paraná), followed by (comma delimited) municipality, followed by (comma delimited) locality (in this case, abbreviations can be used when broadly known and used; e.g. P.E. standing for Parque Estadual/State Park), followed by (comma delimited) the date, followed by (comma delimited) coordinates, followed by (comma delimited) elevation (recorded in metres above sea level; can be documented as elevation range; frequently, this information is not documented), followed by (tab delimited) name of the collector(s), with the first and middle names abbreviated (e.g. R. M. Feitosa standing for Rodrigo Machado Feitosa; in some cases, the surname of the collector precedes the initials of the first and middle names, e.g. Feitosa R. M.), followed by (tab delimited) the sampling technique (sometimes this information is not documented). Instead of the sampling technique, sometimes a code can be used to document sampling approaches and techniques (e.g. P1T1, standing for Pitfall 1 and Transect 1).

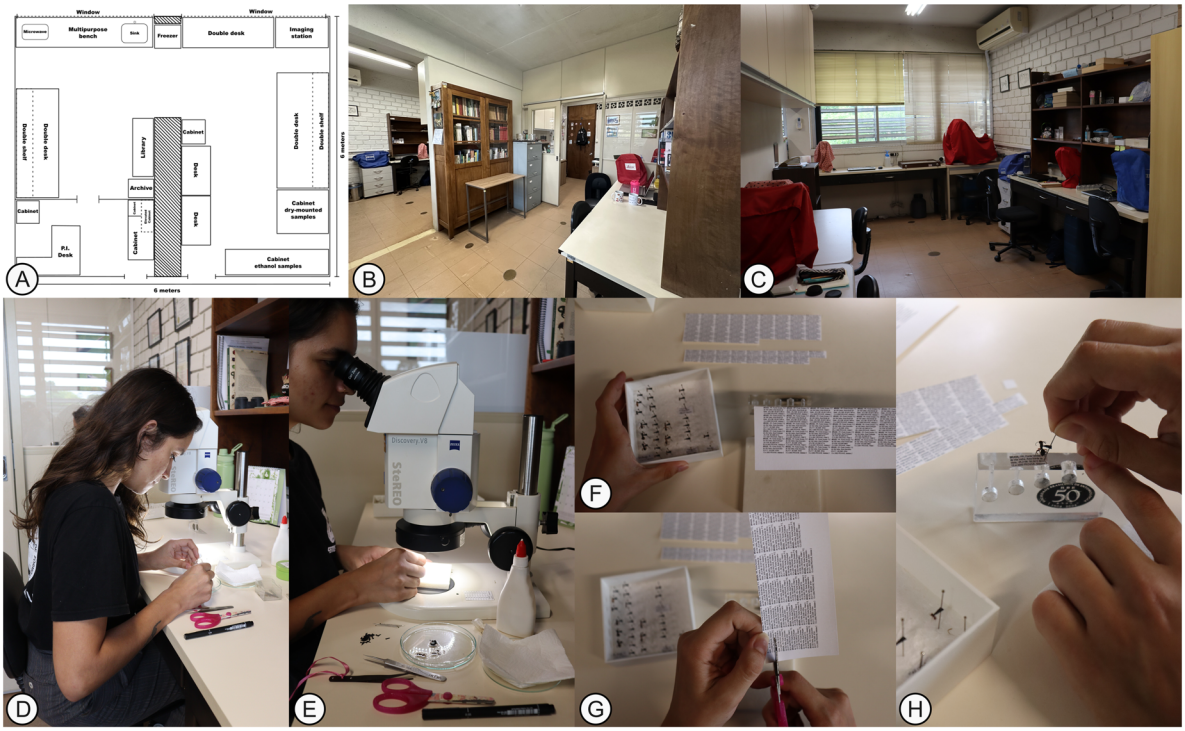

Figure 3. Laboratory infrastructure and personnel, illustrating the process and the places where sample processing and permanent label preparation occur. A, LSBF floor plan. B, north–northwest wing of the laboratory, viewed from the upper left corner of the room. C, south–southeast wing of the laboratory, viewed from the bottom left corner of the room. D, specialist initializing sample-processing procedures. E, specialist preparing a specimen for dry-mounted preservation. F, G, association between dry-mounted specimens and permanent labels. H, dry-mounted specimen associated with a permanent label. Floor plan and image B by Thiago Silva; images C–H by Simone Simioni. Abbreviation: LSBF, Laboratório de Sistemática e Biologia de Formigas.

In particular to dates, some attention has to be taken to avoid errors during transcription: (i) months are usually written in Roman numerals, avoiding confusion during reading, considering that in some countries it is common to transcribe the month before the day (i.e. month/day/year instead of day/month/year); and (ii) the year must be written in four digits to avoid confusion about the year in which the collection event took place (e.g. use 2024 instead of 24). If additional information deemed relevant to documentation exists (e.g. 'specimens collected in a decomposing trunk'), a separate label is prepared to include the transcribed relevant information.

In particular to coordinates, many different coordinate notation formats can be used, depending on the usage history of the research group or on the equipment used to obtain the coordinates. Frequently used formats are degree based ('decimal degrees', 'degree minutes', and 'degree minute seconds') and the Universal Transverse Mercator (UTM) notation. Sometimes, coordinates are omitted from provisional labels influenced by the unavailability of equipment when documenting the collection event; in other cases, the coordinates are purposefully omitted by field researchers. In both cases, the researcher responsible for preparing the permanent label documents the coordinates of the sampling locality in two possible ways: (i) coordinates are not added, given the uncertainty on the exact locality that the specimen/sample was obtained; or (ii) approximate coordinates are defined using the coordinates of the nearest reference point to the sampling site.

Ideally, global (meta)data standards (e.g. Darwin Core) should guide the creation of permanent specimen labels to facilitate biodiversity data sharing. However, these standards often offer limited utility for curators of local biological collections. For instance, the Darwin Core date format (e.g. 1809-02-12 for 12 February 1809) is not easily interpreted by users of the LSBF collection, which primarily serves undergraduate and graduate students. Instead, formats like day/month/year (e.g. 12 Feb 1809 or 12.ii.1809) are preferred and have been adopted as the local standard, which, in the case of the myrmecological community where the LBSF is located, is widely used in both physical natural history collections and online databases.

Additionally, not all samples that receive permanent labels are digitized immediately and therefore do not receive global unique identifiers (GUIDs). In our case, institutional digitization protocols require a specimen or sample to be digitized before a GUID can be assigned, and access to the database management system is limited to specific staff. As a result, this step is often delayed. In practice, GUIDs are assigned only when the material is incorporated into a study or when a technician is specifically hired for digitization, with the latter occurring infrequently. The Brazilian case thus reflects slow digitization owing to limited resources and staffing, even when careful attention is devoted to the material roles of specimen labels.

Field documentation during collection events plays a central role in shaping how epistemic commitments and scientific norms are transmitted within the LSBF. One explicit purpose of these

activities is to familiarize new members with the standard work routines of the group. Coordinators (in this case, T.S.R.S. and R.M.F.) instruct them on how sites are selected, how preferred sampling procedures are conducted, and which types of information are considered relevant to ongoing research. At the same time, participants are taught implicitly what types of details are regarded as less important and therefore excluded from documentation. Through the preparation of provisional labels, new members engage with these priorities in practice. Whether or not they choose to adopt all of them, participation in the event and the act of materializing epistemic commitments in label form influence how they perceive and perform their role as researchers. In this way, specimen labels also fundamentally craft the professional identities of researchers.

However, this pedagogical role of specimen labels is not universal and depends heavily on the infrastructure that supports them. During T.S.R.S.'s work in Hong Kong, the standard practice was to assign specimens a minimal label containing only a unique identification code, whereas detailed locality and sampling information were recorded in a shared cloud-based spreadsheet. In practice, this system often broke down: specimens were labelled but not linked consistently to spreadsheet entries, sampling information was often omitted or forgotten, and records were sometimes stored only on personal computers, without being uploaded or shared with the principal investigator. This contrasts with the Brazilian case, where provisionality is invested with pedagogical value and later stabilized into standardized permanent labels. In Hong Kong, provisionality is compressed into a minimal placeholder, reliant on digital infrastructures that were maintained inconsistently.

These inconsistencies reflected broader institutional priorities. In Hong Kong, funding schemes privileged species records and taxonomic novelties over comprehensive documentation of collection events, a focus that also shaped policies for collection management. Acquisition of materials for documentation, such as specialized paper for specimen labels, was constrained by policies requiring purchases from registered suppliers, and reimbursement processes were slowed by bureaucratic procedures. Infrastructure for housing collections received little institutional support, whereas projects involving extensive field sampling were readily funded. In Brazil, in contrast, material care and permanent labelling were strongly emphasized, but digitization was delayed and precariously funded. Together, these cases reveal how local practices diverge in their relationship to global agendas: Brazil highlights adaptation and delay, whereas Hong Kong illustrates acceleration and fragmentation.

These contrasting scenarios illustrate how global agendas for digitization reshape local practices. Institutions often channel resources towards meeting international targets for data accessibility, funding digitization projects while overlooking the material and curatorial work of maintaining physical labels and collections. As a consequence, the efforts of researchers in producing and caring for specimen labels are perceived as peripheral to the demands of global data infrastructures. The undervaluing of this labour not only undermines local epistemic priorities but also affects how the contributions of researchers are recognized within their communities. Institutional hiring practices reinforce this imbalance: technicians tasked with digitization are more readily employed than archivists, curators, museologists, or taxonomists responsible for long-term care of biological collections, and such technician positions are frequently underfunded and precarious. For example, fixed-term contracts for digitizing entomological collections in Brazil can require doctoral-level expertise in taxonomy and curation yet offer modest remuneration (approximately USD 460 per month for a 40 h working week). In Hong Kong, resource allocation favoured research outputs, whereas bureaucratic procurement processes constrained investments in material curation. Despite their differences, both cases demonstrate how short-term digitization goals are prioritized over sustained investment in the preservation and epistemic significance of biological collections.

Record entanglement: materiality and infrastructural linkage of specimen labels in biological collections

One of the greatest conflicts that researchers face when working with biological collections concerns the materiality of specimen labels. Frequently, this materiality is seen as cumbersome in the daily routine of a laboratory or biological collection owing to two main factors: (i) its elaboration demands time; and (ii) its presence demands space. Unfortunately, in contemporary scientific making, lack of time and lack of space are frequently observed, especially in institutions spatialized in the so-called Global South: (i) researchers need to analyse a huge amount of data in a short time in order that they can publish more (and in 'higher-quality' journals); and (ii) the physical infrastructure is limited, because funding for creation and maintenance of collections is insufficient and the availability of space for storage of materials is limited. Whether researchers like it or not, 'dealing' with the materiality of a specimen label demands time and requires specialized resources (Hopman 2024).

Beyond logistical concerns, materiality intersects with meaning. As Keane (2005) observes, structural assumptions often divide focus between 'things' and 'ideas': emphasizing one tends to reduce the other to an epiphenomenal role. In the case of specimen labels, this results in a tendency to treat the label as a mere container of data, i.e. merely recording the occurrence of a specimen at a location, and assigning this record to a species. Although this description is not entirely inaccurate, it overlooks the complex practices through which labels are produced in the field and curated in collections.

Fieldwork demonstrates that the process of labelling is shaped by interactions between individual preferences and collective protocols. For instance, during collection events at the Parque Estadual das Lauráceas, T.S.R.S. observed that collectors often abbreviated locations (e.g. 'Cemitério' instead of 'Parque Estadual das Lauráceas—Trilha do Cemitério') to save time, which later required laboratory recorders to cross-check field notebooks or consult with the principal investigator during permanent label preparation and/or digitization. Likewise, generalized location labels (e.g. 'Parque das Lauráceas Sede') sometimes obscure micro-habitats important for ecological analysis. These examples show that recording practices are embedded in specific institutional and field contexts, and the frictions between individual choices and collective standards directly shape the material form of labels. These frictions do not remain confined to the field but travel with the specimen labels into subsequent stages of curation and digitization, where vague or inconsistent information can

complicate cataloguing, retrieval, taxonomic studies, and ecological analysis, often requiring additional verification or reconstruction of provenance. Such complications are not only technical but also reputational, because the perceived quality of documentation becomes associated with those responsible for producing or curating the labels, influencing how their scientific work is judged.

Hence, the influence of these practices extends to the routines and professional identities of researchers. Micro-routines of collection, curation, and laboratory work orient researchers to particular spaces and object forms, which, in turn, shape expectations and habits. Recorders develop reputations for thorough, sparse, or inconsistent documentation, influencing the perceived quality and, by extension, the reliability of their scientific outputs [e.g. incorrect type locality information in Forel's protonyms, highlighting discrepancies between specimen labels and catalogue records (Ulysséa *et al.* 2025)]. In this sense, material forms of labels become intertwined with the professional identities of researchers (Bourdieu 1977).

At the same time, although specimen labels document scientific activity, they often obscure the contributions of other agents. Technicians, cleaners, electricians, and artisans involved in maintaining collections are frequently excluded from labelling records and the environments they sustain. This exclusion has practical consequences: the omission of provenance or maintenance information can lead to repeated errors, loss of contextual knowledge, and under-recognition of the labour required to uphold collections.

Specimen labels, in combination with other non-prototypical records (e.g. field notes, maps, collection logs), form an integrated system that supports the creation and retrieval of knowledge (Griesemer 1990). These heterogeneous records constitute an infrastructure that underpins scientific practice: they occupy an intermediate space between material reality and scientific facts, mediating between data and formulated knowledge (Shankar 2009). Observations made by Ulysséa *et al.* (2025) in the ant collection of the Museu de Zoologia da USP (MZSP) show that access to collection logs, field notes, and specimen labels is crucial for researchers' decision-making and understanding of specimens, demonstrating that infrastructure is not merely background but actively shapes the production of knowledge.

Ignoring this entanglement of records risks erasure of the processes through which scientific knowledge is produced, concealing both academic and non-academic contributions, and masking the procedural, epistemic, and ethical dimensions of research. For example, decisions about which species to document or which local knowledge to include can reflect both biological biases and social inequalities, with direct consequences for conservation priorities and the management of collections (Halm and Santana 2025).

DISCUSSION

Reframing specimen labels from 'data chunks' to records of activities is important because we start comprehending the process of creating knowledge and making decisions in conservation beyond the anxious need for data extraction and the constant 'doom-scrolling' in databases. Although the increase in the volume of data is seen as important to orient decision-makers and stakeholders in biodiversity conservation matters, exposing the entanglements among the elements that constitute the infrastructure of scientific making is fundamental to orient governmental and non-governmental institutions on ways to support and preserve it.

Data do not exist by default; they are construed and reflect the dynamics of several agents and dimensions, including the spatialization of those same agents. It is extremely problematic when we focus on increasing data availability and 'accessibility' without considering the infrastructure that produces and maintains them; this approach not only commodifies agents and their daily life struggles but also maintains and legitimizes hierarchies of scientific making, with handling and care of material entities being subordinated to abstractions of causality and universality. As illustrated by Hopman (2024) in her ethnographic study of the digitization of the Museum für Naturkunde Berlin's collection of snail shells, the process of mass digitization works by an expansive logic similar to anticipatory regimes (Adams *et al.* 2009). According to Adams and her colleagues, anticipatory regimes (such as those of capitalism) tend to work through logics of expansion, in which new territories for speculation must continually be found to keep the anticipatory logic moving; in this sense, anticipatory regimes expand their scope of inclusion and elongate their reach in time, in space, and in phenomenological terms (Adams *et al.* 2009).

The manufacture of 'global natural products', in the form of atomized data about biodiversity, strips knowledge produced by marginalized agents and institutions, erases the histories, procedures, and capacities directly or indirectly involved in the activities of specimen label preparation, and repackages it as privatized commodities. These 'resources' ultimately serve the interests of Intergovernmental Panels that dictate 'global' conservation frameworks and goals, reinforcing an ethnocentric, extractivist view of natural and historical heritage. This process upholds colonial governance of 'nature', positioning these panels as the self-appointed custodians of universal knowledge, hoarding it within their servers while silencing the very communities whose expertise and capacities sustain it.

A consequence of this process is the deterioration of public institutions and their funding policies in places spatialized in the so-called Global South. One example is the dissolution and extinction of the Fundação Zoobotânica do Rio Grande do Sul (FZB), one of Brazil's most respected environmental institutions, which housed the Museu Riograndense de Ciências Naturais, the Jardim Botânico de Porto Alegre, and the Parque Zoológico de Sapucaia. All entities previously linked to the FZB are now abandoned, and, as a result, the agents and capacities associated with them have been scattered, because the private sector has shown little interest in funding them. A similar pattern can be seen in the weakening of the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) in Mexico, which has been downgraded from a multi-ministerial federal agency to a division within the Secretaría de Medio Ambiente y Recursos Naturales (Medellin and Soberón 2024).

As highlighted by Brandão *et al.* (2021), documentation in biological collections is often neglected, and good practices for the management of biological collections, whether national or institutional, should be accompanied by appropriate documentation measures. We argue further that policies governing biological

collections should establish clear guidelines for curating specimen labels and all other non-prototypical records (e.g. field notebooks, illustrations, photographs, audio recordings). Pragmatically, national policies (such as Brazil's PL 1993/2024 bill on biological collections) must recognize explicitly that these collections encompass more than biological specimens alone; they also include diverse non-biological materials.

Moreover, national policies for biological collections often focus on ensuring the security, accessibility, quality, longevity, integrity, and interoperability of biodiversity data, aligning with global initiatives for open linked data [e.g. FAIR and the Global Biodiversity Information Facility (GBIF)], yet fail to provide comparable guidelines for non-specimen materials (e.g. labels, field notes, or multimedia records). This oversight not only leaves most non-prototypical records inadequately managed but also hinders engagement with marginalized stakeholders, including local educators (from kindergarten to high school) and community members. Without clear policies governing access to specimen labels and other types of non-prototypical records, biological collections miss crucial opportunities to foster inclusive dialogue and public participation.

In terms of providing visibility to the agents involved in maintaining the infrastructure of a biological collection, the use of universal identifiers for researchers and research institutions (i.e. ORCID) in the documentation of biological collections could be a possibility.

However, most of the agents involved in maintaining the infrastructure of a biological collection are not necessarily embedded in the 'scientific culture' as researchers (e.g. technicians), and they do not necessarily have to be. The use of national universal identifiers (e.g. the *Brazilian Cadastro de Pessoa Física*) for documentation of agents involved in biological collections would be preferred, because those agents not involved in scientific research would still gain visibility and have their sense of belonging manifested as formal connections to the biological collection via institutional contract. However, the use of such identifiers must be balanced carefully with data protection and privacy considerations (in our example, particularly under Brazil's General Data Protection Law, LGPD), to ensure that personal information is handled responsibly and with appropriate consent.

Exposing the different dimensions involved in specimen label production is not only important for scientific development; it serves a broader purpose. Analysing this 'forgotten' infrastructure of the life sciences is necessary for the survival of these 'subaltern' elements (non-prototypical records, agents, institutions, social dynamics, learning, and teaching) inscribed in this 'hierarchy of valuable scientific knowledge'.

Pragmatically, we suggest two immediate solutions: (i) dispel the imaginary of global digitization initiatives as an equalizing power in biodiversity collection management and understand how these initiatives affect funding policies at finer scales; and (ii) expose national funding deficiencies for the management of biological collections via the study of access to and preservation of non-prototypical records.

Instead of 'keep(ing) calm and digitiz(ing) everything', we should, along with digitization strategies, expose and document the frictions and struggles of the agents and institutions involved in the creation and preservation of the artefacts being digitized. Documenting the history of specimen labels allows us to understand all the power struggles that are pervasive in this data production line. It reveals how local activities and capacities are diluted, and eventually erased, in the final commodified product.

AUTHOR CONTRIBUTIONS

T.S.R.S. proposed the idea and the project. T.S.R.S. and R.M.F. performed sample processing on the field and fieldwork documentation. A.C.A.N. performed sample processing on the laboratory and laboratory documentation. A.C.A.N. and R.M.F. supervised the project. T.S.R.S. and A.C.A.N. conceptualized and structured the paper. T.S.R.S. and A.C.A.N. wrote preliminary versions of the manuscript. All authors discussed the results, commented on the manuscript, and wrote the paper.

CONFLICT OF INTEREST

None declared.

FUNDING

T.S.R.S. was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–CAPES, under the program CAPES/PRINT (grant number 88887.892274/2023-00). R.M.F. was supported by a National Council for Scientific and Technological Development (CNPq) Productivity Grant (304012/2023-8).

DATA AVAILABILITY STATEMENT

The data underlying this article are all available in the article's text.

REFERENCES

Adams V, Murphy M, Clarke AE. Anticipation: technoscience, life, affect, temporality. *Subjectivity* 2009;**28**:246–65.

Barad K. *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham and London: Duke University Press, 2007.

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

Berkes F. From community-based resource management to complex systems: the scale issue and marine commons. *Ecology and Society* 2006;**11**:45.

Bourdieu P. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press, 1977.

Bowker GC. Biodiversity Datadiversity. *Social Studies of Science* 2000;**30**:643–83.

Brandão CRF, Ramos KDS, Ulysséa MA *et al.* Princípios para a curadoria técnica do acervo entomológico do Museu de Zoologia da Universidade de São Paulo. *Anais do Museu Paulista: História e Cultura Material* 2021;**29**:e31.

Corrado AR, Gagliatti AL, Romaniuc Neto S *et al.* Reports of the use of Urticaceae collected in Brazil and deposited in the herbaria of Kew (K), New York (NY) and Paris (P). *Ethnobiology and Conservation* 2015;**4**:1–12.

Gitelman L, Jackson V. Introduction. In: Gitelman L (ed.), *"Raw Data" is an Oxymoron*. Cambridge and London: MIT Press, 2013, 1–15.

Graham CH, Ferrier S, Huettman F *et al.* New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* 2004;**19**:497–503.
Griesemer J. Modeling in the museum: on the role of remnant models in the work of Joseph Grinnell. *Biology and Philosophy* 1990;**5**:3–36.
Halm D, Santana C. Debiasing collection in field biology. *Philosophy, Theory, and Practice in Biology* 2025;**17**:5.
Hardisty AR, Ellwood ER, Nelson G *et al.* Digital extended specimens: enabling an extensible network of biodiversity data records as integrated digital objects on the internet. *Bioscience* 2022;**72**:978–87.
Holetschek J, Dröge G, Güntsch A *et al.* The ABCD of primary biodiversity data access. *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology* 2012;**146**:771–9.
Hopman R. Snails, time, data: on the politics of mass-digitization and the possibility of data drift. *Big Data & Society* 2024;**11**:20539517241267760.
Horning NR, Bridging the gap between environmental decision-makers in Madagascar. In: German LA, Karsenty A, Tiani A-M (eds.), *Governing Africa's Forests in a Globalized World*. London and Virginia: Earthscan Press, 2010, 234–57.
Ilerbaig J. Specimens as records: scientific practice and recordkeeping in natural history research. *The American Archivist* 2010;**73**:463–82.
Johnson KR, Owens IFP; Global Collection Group. A global approach for natural history museum collections: integration of the world's natural history collections can provide a resource for decision-makers. *Science (New York, N.Y.)* 2023;**379**:1192–4.
Keane W. Signs are not the garb of meaning: on the social analysis of material things. In: Miller D (ed.), *Materiality*. Durham and London: Duke University Press, 2005, 182–205.
Medellin RA, Soberón J. Don't bury Mexico's biodiversity capacity. *Science (New York, N.Y.)* 2024;**384**:9.
Pearce SM. *Museum Studies in Material Culture*. London: Leicester University Press, 1989.
Rudolph EA, Wiman NG. Insights from specimen data for two economic *Chrysobothris* species (Coleoptera: Buprestidae) in the western United States. *Annals of the Entomological Society of America* 2023;**116**:195–206.
Schreiner K. *Terminological Dictionary of Museology*. Berlin: s. n., 1988.
Shankar K. Ambiguity and legitimate peripheral participation in the creation of scientific documents. *Journal of Documentation* 2009;**65**:151–65.
Stewart R. *Request for Comments (RFC) 4960 Stream Control Transmission Protocol (SCTP)*, 2007. Available from https://datatracker.ietf.org/doc/html/rfc4960 (29 November 2024, date last accessed).
Thessen AE, Cui H, Mozzherin D. Applications of natural language processing in biodiversity science. *Advances in Bioinformatics* 2012;**2012**:391574.
Ulysséa MA, Santos S, Moleiro HP *et al.* The MZSP Ant Type Collection: a complete catalogue and digital resource for global access. *Biological Journal of the Linnean Society* 2025;**145**:blaf034.
van Mensch P. Methodological museology; or, towards a theory of museum practice. In: Pearce SM (ed.), *Objects of Knowledge*. London and Atlantic Highlands: Athlone Press, 1990, 141–57.
Vollmar A, Macklin JA, Ford L. Natural history specimen digitization: challenges and concerns. *Biodiversity Informatics* 2010;**7**:93–112.
Wood N. Archive and library special collections as proxy data: reconstructing the American chestnut blight through digitized collections. *Archival Science* 2025;**25**:4.
Yeo G. Concepts of record (1): evidence, information, and persistent representations. *The American Archivist* 2007;**70**:315–43.
Yeo G. Concepts of record (2): prototypes and boundary objects. *The American Archivist* 2008;**71**:118–43.