Received: 10 February 2025 | Revised: 9 May 2025 | Accepted: 22 May 2025
DOI: 10.1002/ppp3.70052

METHODS AND TECHNIQUES

Plants People Planet PPP Open Access

Repositories of biocultural diversity: Toward best practices for empowering ethnobotany in digital herbaria

Robbie Hart[1] | Viviane Fonseca-Kruel[2] | Eduardo Dalcin[2] | Estevão da Silva Luís Alexandre[2] | Matthew Pace[3] | Michaela Schmull[4] | Leonardo Beltran-Rodríguez[5] | Miguel Murguía-Romero[6] | Diana G. Flores-Camargo[5] | Cristina Mapes-Sánchez[5] | Mark Nesbitt[7] | Carolina Romero[1] | Andrew Townesmith[1] | Jan Salick[1] | Alex McAlvay[3] | Kenneth R. Otero-Walker[3] | Michael J. Balick[3] | Jacob Golan[8] | Bruce Hoffman[9] | Kelsey Leonard[10] | Giulia Mattalia[11] | Guillaume Odonne[12] | Aurora Prehn[1,7,13] | Ina Vandebroek[14]

[1]William L. Brown Center, Missouri Botanical Garden, Saint Louis, Missouri, USA

[2]Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil

[3]Center for Plants, People and Culture, New York Botanical Garden, New York, New York State, USA

[4]Harvard University, Cambridge, Massachusetts, USA

[5]Jardín Botánico, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

[6]Unidad de Informática para la Biodiversidad, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

[7]Science, Royal Botanic Gardens, Kew, London, UK

[8]New York University, New York, New York State, USA

[9]Amazon Conservation Team, Falls Church, Virginia, USA

[10]University of Waterloo, Waterloo, Canada

[11]IMBE, Aix Marseille Université, Avignon Université, CNRS, IRD, Marseille, Bouches-du-Rhone, France

[12]LEEISA-UAR3456, CNRS, Université de Guyane, IFREMER, Cayenne, France Guiana, France

Societal Impact Statement

As herbaria digitize millions of plant specimens, ethnobotanical information associated with them is becoming increasingly accessible. These biocultural data include plant uses, names, and/or management practices of Indigenous Peoples and Local Communities (IPLCs). However, the absence of shared curatorial standards limits accessibility and use by IPLCs and others. We estimated and characterized ethnobotanical data associated with herbarium specimens and provide here key considerations for future work. We identified a proportionally small, yet collectively significant, number of ethnobotanical specimens, and call for coordinating best practices among global herbaria to locate, acknowledge, and responsibly share this information, together with source communities.

Summary

- As herbaria digitize millions of plant specimens, those containing biocultural information are becoming increasingly accessible. This information — also known as ethnobotanical data — holds both cultural and scientific value, and may include plant uses, vernacular names, local species concepts, cultural values, and plant management practices of Indigenous Peoples and Local Communities (IPLCs). However, the lack of coordinated curatorial standards currently limits both the accessibility and effective use of this information by IPLCs, ethnobotanists, and others.
- To address this gap, we quantitatively estimated and characterized ethnobotanical information associated with herbarium specimens and offer key considerations to guide future work.

[13]Department of Geography, Royal Holloway, University of London, Egham, Surrey, UK

[14]Caribbean Centre for Research in Bioscience (CCRIB) and Department of Life Sciences, Faculty of Science and Technology, The University of the West Indies, Mona, Kingston, St Andrew, Jamaica

Correspondence
Aurora Prehn, William L. Brown Center, Missouri Botanical Garden, Saint Louis, Missouri, USA.
Email: aprehn@mobot.org

- We identified a proportionally small —yet collectively significant— number of ethnobotanical specimens, comprising approximately 1.6% of all specimen records and representing hundreds of thousands of specimens in the surveyed herbaria.
- We advocate for coordinating best practices to locate, acknowledge, and ethically share this information among herbaria, working together with source communities and through global cooperation.

KEYWORDS
biocultural collections, data sovereignty, digitization, ethnobiology, indigenous and local knowledge, languages, plant taxonomy, traditional ecological knowledge

1 | INTRODUCTION

Globally, an estimated 3,864 active herbaria house collections of preserved plant specimens that support botanical research by providing critical information for taxonomy, systematics, ecology, and biodiversity conservation (Bebber et al., 2010; Primack et al., 2004; Thiers, 2020; Thiers, 2025). Beyond their main purpose, specimen labels may also contain biocultural information (ethnobotanical data), such as plant uses, vernacular names, traditional ecological knowledge, and market prices (Salick et al., 2014). Ethnobotanical data held within biocultural collections offer valuable insights into the deep-rooted relationships between plants and people, including cultural and spiritual values, stewardship practices, and knowledge systems. These data can inform biodiversity conservation efforts, support cultural resilience, be useful to develop outreach initiatives together with source communities and the broader public, and contribute to the United Nations Sustainable Development Goals (Arrivabene et al., 2024). Nevertheless, the full potential of herbaria to effectively store, analyze, and ethically share these ethnobotanical data and their associated biocultural collections has yet to be realized.

Early efforts to retrieve ethnobotanical data from specimen labels relied on manual consultation of non-digitized collections. This labor-intensive process involved examining each herbarium sheet. Half a century ago, botanist Siri von Reis used this method to review field notes from about 2.5 million globally collected specimens and recorded medicinal uses for over 5,000 species (von Reis, 1973). This and later studies (von Reis & Lipp, 1982) highlight the latent repository of biocultural information contained in herbaria. With an estimated 350–400 million botanical specimens around the world (Delves et al., 2024; Thiers, 2025), the global scale of dormant ethnobotanical data is significant and requires updated practices for data collection, documentation, digitization, access, governance, sharing, and care.

Moreover, the direct association between ethnobotanical data and botanical data linked to voucher specimens ensures verifiable plant identification that is anchored in specific times, geographic locations, ecologies, and cultural contexts (Nesbitt, 2014; Odonne et al., 2021; Vandebroek et al., 2018). These high-resolution, specimen-linked data offer exciting opportunities for research, conservation, and the support of cultural resilience and human wellbeing. Digitized herbarium specimens, which provide records of plant use at specific times and places, enable the global academic community to conduct more precise analyses of human-plant relationships. Furthermore, accessible specimens can support the rights of Indigenous Peoples and Local Communities (IPLCs) by offering greater cultural context and details on knowledge sourcing and attribution, which are important considerations in ongoing discussions of intellectual property rights and digital genetic sequence information (Cowell et al., 2021).

1.1 | Challenges to integrating ethnobotanical data from herbaria

Despite their potential as repositories of ethnobotanical data and biocultural specimens, herbaria face significant challenges in effectively incorporating, managing, and sharing these data.

Firstly, herbaria often contend with financial limitations, limited staffing, and a primary focus on education and physical upkeep (Delves et al., 2024). These constraints can hinder herbaria's ability to allocate sufficient time and resources for digitization efforts. While the use of digital herbarium data has grown significantly in recent years, data aggregation in platforms, such as the Global Biodiversity Information Facility (GBIF), tends to prioritize certain types of information and certain institutions. For instance, although GBIF includes contributions from institutions across the Global North–South divide, those in Europe and the United States still provide the majority of specimen data (Delves et al., 2024).

Secondly, several barriers hinder the systematic recording of ethnobotanical data during plant collection efforts and their subsequent curation. Historical biases in plant collecting have favored geographic areas and species of primary interest to specialized taxonomists, often resulting in the uneven documentation of ethnobotanical data. Lack of searchable ethnobotanical data in herbaria limits their accessibility and usability. The Economic Botany Data Standard (Cook, 1995) represented an early effort to standardize ethnobotanical data collection. However, its global applicability is constrained by the fact that different institutions may have different capacities and objectives. As a result, Kuhlman and Salick (2014) recommended inter-institutional compatibility of databases rather than strict standards. For instance, ethnobotanical use categories in the Tropicos database used by Missouri Botanical Garden were designed to be compatible with, though simplified from, the Economic Botany Data Standard (Kuhlman & Salick, 2014). Additionally, even specimens that contain rich biocultural information (e.g., those collected as part of ethnobotanical studies) may become disconnected from the agreements, methods, and sociocultural contexts that informed their collection. When scholars publish these, a best practice would involve citing or linking the specimen (or its digitized metadata) in the publication and also referencing the publication on the specimen record. This bidirectional linkage can facilitate accurate and contextualized interpretation and use of biocultural specimens.

Thirdly, ethical use and sharing of ethnobotanical data from herbarium specimen labels—particularly traditional knowledge from IPLCs—requires careful attention (Levis et al., 2024). Ethnobiologists have long advocated for practices that respect IPLCs' intellectual property contributions, ensure fair benefit-sharing, and adhere to international agreements such as the 1992 Convention on Biological Diversity and the Nagoya Protocol (Secretariat of the Convention on Biological Diversity, 1992). Herbaria generally recognize the importance of balancing information accessibility with the need for certain protections, for instance, in masking the geographic locations of some threatened species. Extending such tools to ethnobotanical data is possible but has yet to be implemented widely or in a coordinated manner. The International Society of Ethnobiology's Code of Ethics (2006) offers comprehensive guidelines for conducting ethical fieldwork and managing ethnobotanical data, but they have been applied mainly within the disciplines of ethnobotany and ethnobiology. Complementary principles such as F.A.I.R. (Findable, Accessible, Interoperable, Reusable) and C.A.R.E. (Collective Benefit, Authority to Control, Responsibility, Ethics) are also available to guide actionable, equitable, and inclusive data practices (Carroll et al., 2020; Davis et al., 2025; Wilkinson et al., 2016).

1.2 | Hypothesis and aims

Our central hypothesis is that herbarium collections contain a substantial number of specimens with biocultural information whose utility and visibility are limited by incomplete or incompatible recording formats. We propose that enhancing data compatibility would facilitate more effective cross-institutional comparisons, support global-scale analyses, and improve accessibility for users, thereby increasing the recognition and use of this data.

In this context, the specific objectives of this paper are to:

1. Quantify the presence and scope of ethnobotanical data within selected digitized herbarium collections, using collectively agreed upon database search keywords;
2. Highlight the challenges inherent to retrieving and comparing ethnobotanical data across institutions that differ in size, geographic context, and curatorial history, and compare the results of our joint keyword searches with those of existing institutional projects
3. Demonstrate the need for the future development of a coordinated framework of best practices to guide the inclusion, access to, and ethical sharing of ethnobotanical data in herbaria.

2 | METHODOLOGY

Prior to this study, the authors formed an ad hoc working group that convened regularly via Zoom to discuss the biocultural collections within their respective institutions. This effort was built upon earlier initiatives (Salick et al., 2014), including workshops held during annual meetings of the Society for Ethnobotany (formerly the Society for Economic Botany) under NSF award #1118808. Previously, several authors had already independently undertaken projects using various methods to locate, sample, or summarize herbarium specimens with ethnobotanical data (e.g. Flores-Camargo & Sánchez-Dirzo, 2022). Group discussions about these efforts, which revealed shared concerns in identifying and making ethnobotanical data accessible, were a key part of our methodology and informed our results and recommendations. Recognizing that herbaria hold overlapping yet largely distinct collections, and operate within similar frameworks, the group agreed on the need for developing a set of guiding principles for advancing our specific objectives.

As an initial step to quantifying the scope of biocultural information, we collaboratively implemented a keyword text search of digitized herbarium records from five institutional databases to retrieve specimens of ethnobotanical relevance. Keywords were selected by consensus to encompass diverse plant use categories (e.g., medicines, foods) while accounting for the linguistic diversity of specimen labels (Table 1). By comparing data across herbaria, we aimed to identify patterns in plant use information, gaps, and inconsistencies, highlight the methodological challenges inherent in such a comparative analysis, and suggest steps toward a more refined methodology.

Despite the variety of contexts and software used, the implementation of our methodology revealed several recurring challenges. Primarily, the absence of (or lack of data in) dedicated ethnobotanical fields necessitated searches within unstructured text fields (e.g., "Description," or Darwin Core's "occurrenceRemarks"). This was further complicated by instances where these unstructured fields were also unpopulated, a consequence of past digitization projects prioritizing structured data. Moreover, even when transcribed, the limited searchability of unstructured text hindered efficient data retrieval and analysis. In some cases, accessing data requires special access privileges, and in most cases, it requires some level of expert knowledge.

Searching transcribed text fields also posed language and software-specific challenges around delimiter characters (such as commas and tabs) which can complicate data conversion, character variations (such as accented characters), and word boundaries (with some but not all software searching within words by default). Institutions with experience in thematic projects involving ethnobotanical herbarium specimens, such as the Maximino project at the National Herbarium of Mexico (MEXU) (Flores-Camargo &

TABLE 1 Included keyword search terms in English, Spanish, Portuguese, and French based on consensus discussions among the authors and reflecting the primary languages represented in their collections. Word fragments reflect inclusive searching within words (e.g. 'medic' returns 'medical', 'medicine', etc.).

Category	Keyword search terms
Medicinal	*English*: Health, medic, to cure, cure for, treat, fever, febrifuge, malaria
	Spanish: Salud, medicin, Para curar, fiebre, febrífugo, malaria
	Portuguese: Saúde, medicin, Para cura, tratamento, febre, febrífugo, malária
	French: Santé, médecine, fièvre, fébrifuge, paludisme
Food	*English*: Food, aliment, crop, edible
	Spanish: Alimento, cultivo, comestible
	Portuguese: Alimento, cultura, comestível
	French: Nourriture, culture, comestible
Spiritual	*English*: Ritual, magic, religion, spiritual
	Spanish: Ritual, mágico, religion, espiritual
	Portuguese: Ritual, mágico, religioso, espiritual
	French: Rituel, magie, religion, spirituel
Other uses	*English*: Fibre, fiber, cordage, string, rope, basket, thatch, roofing, material, timber, construction, dye, dyestuff, colorant, poison
	Spanish: Fibra, cuerda, cordón, canasta, materiales Para el techo, paja Para el techo, Madera, construcción, tintura, colorante, venen
	Portuguese: Fibra, corda, cordão, barbante, cesta, palha, cobertura, construção, corante, tintura, pigmento, veneno
	French: Fibre, cordage, ficelle, corde, panier, chaume, toiture, matériau, bois, construction, colorant, poison
Vernacular name	*English*: Vernacular, n.v.
	Spanish: Nombre común, n.c.
	Portuguese: Nome vernacular, Nome comum, Nome popular
	French: Nom vernaculaire

Sánchez-Dirzo, 2022), have identified text string searches using regular expressions (sequences of characters used to find, match, and manipulate character strings that allow highly customizable searches, Friedl, 2006) as an effective tool for processing search results. This approach enables the use of a broad range of search terms while minimizing 'bycatch' (false positives) where a search term appears within another context that does not indicate biocultural information.

3 | RESULTS

3.1 | Keyword search results across herbaria

Our keyword searches examined 14.8 million specimens from 104 herbaria and identified 231,136 specimens with potential ethnobotanical data (about 1.6% of the total, see Table 2). The proportion of these retrieved specimens varied considerably among the participating herbaria, ranging from 0.4% to 2.6%.

We acknowledge that the results in Table 2 involve both under- and overcounting. Expanding the number of keyword search terms would likely increase the proportion of retrieved specimens, although this effect is expected to plateau as the total number of retrievable specimens is approached. Conversely, reviewing and validating specimens reduces the proportion by filtering out false positives, which would similarly stabilize upon reaching the true number of specimens with ethnobotanical data. Additionally, systematic differences in database structures and methodological variations affect the results. For example, MEXU conducted searches exclusively with Spanish-language terms, JABOT used only Portuguese terms, and NY employed two separate searches: one for fully transcribed records and another for optical character recognition-only records without full transcription. Therefore, the results in Table 2 should be regarded as indicative for estimating the overall scope.

Nevertheless, when individual institutions either expanded or customized their search terms and/or manually reviewed subsets of data to remove bycatch, proportions of biocultural specimens were within or near the range reported in Table 2. For instance, 8,400 recently collected MEXU specimens were searched with an expanded set of locally specific keywords, yielding almost 2,000 records. However, manual review eliminated many of these as bycatch and 263 specimens (3.1%) were validated as containing ethnobotanical data. At MO, a large, customized set of search terms yielded 900,000+ specimens as an initial result, of which many were identified as bycatch with regular expressions, leading to ~100,000 validations (2.4% of the specimens searched). At Kew, scientists working with a subset of digitized specimens and internal keywords returned ~6,700 out of 300,000 specimen records (2.2%). These findings increase our confidence that an estimate of 1.6% is conservatively indicative of the scope of ethnobotanical data in herbaria.

3.2 | Validation and refinement of search results

Automated methods can enhance ethnobotanical data retrieval from herbaria. For example, NY herbarium's use of optical character recognition in searches for terms such as 'community' and 'Indian' uncovered previously overlooked specimens. However, language and specific keyword searches influenced the rate of bycatch. For instance, we noted that 'culture' appears in many other contexts ('agriculture', 'fungal culture', etc.). We originally included the word 'Indigenous', but omitted it given the large number of specimens yielding results related to their geographic range rather than Indigenous knowledge. MO excluded "sante" and "fiber" from the analysis due to high bycatch – in the case of "fiber," tens of thousands of records referred to plant anatomy rather than use.

However, although bycatch was common, these false positives were not the majority of records. For instance, a manual review of

TABLE 2 Results of digitized herbarium collection searches to assess the extent of biocultural specimens, based on search terms from Table 1. Database platforms included EMu (NY, the New York Botanical Garden), Tropicos (filtered to specimens from MO, Missouri Botanical Garden), HUH-specify (HUH, Harvard University Herbaria), IBdata and Maximino (MEXU, National Herbarium of Mexico, Murguía-Romero et al., 2024), and JABOT (100 Brazilian herbaria, note S1). JABOT is a herbarium information system developed by RB, the Rio de Janeiro Botanical Garden, and offered as a "Software for Service" to 100 herbaria in Brazil (Silva et al., 2017). Herbarium acronyms follow Thiers (2025). HUH includes A (1872), AMES (1899), ECON (1858), FH (1919), GH (1868), and NEBC (1896, filed with Harvard's collection but not property of Harvard).

Herbarium acronym (country, founding date) - size in terms of total # specimens	Total # digitized & transcribed specimens searched	# specimens identified that returned results with keywords (% of digitized total)	# taxa (unique genus+species string)
MEXU (Mexico, 1888) - 1.6 million	1.2 million	30,846 (2.6%)	4,605
NY (USA, 1891) - 7.8 million*	4.7 million	37,242 (0.8%)	17,632 (1.5% of taxa in NY database)
MO (USA, 1859) - 8+ million*	4.1 million	89,571 (2.2%)	27,027
HUH (USA, 1858–1919) - 5.5 million*	1.8 million	7,176 (0.4%)	Unknown
JABOT (100 Brazilian herbaria, note S1)–4.1 million	3.0 million	66,301 (2.2%)	15,129
TOTAL: 27+ million	**14.8 million**	**231,136 (1.6%)**	

*Estimated number.

TABLE 3 Subtotals of digitized and transcribed specimens per herbarium derived from keyword search terms in Table 1. Categories include specimens with vernacular names, as well as medicinal, food, spiritual, and other uses. OCR: optical character recognition, a technology used to convert printed or handwritten text on specimen labels into machine-readable text. MEXU, National Herbarium of Mexico; NY, the New York Botanical Garden; MO, Missouri Botanical Garden; HUH, Harvard University Herbaria; JABOT, 100 Brazilian herbaria, listed in note S1.

Herbarium name	# medicinal specimens	# food specimens	# spiritual specimens	# vernacular name specimens	# other use specimens
MEXU	7,553	4,066	47	19,996	1,471
NY (based on transcribed data)	6,470	2,008	2	66,962	18,076
NY (based on OCR searches)	1,389	4,932	152	24,212	6,555
MO	11,182	18,564	236	58,258	1,587
HUH	410	372	11	3,868	2,515
JABOT	5,814	4,750	166	55,123	448

randomly sampled 100 records from the 89,571 records identified at MO identified 70 records that contained biocultural information versus 30 bycatch.

3.3 | Category-specific insights: vernacular names and plant uses

Vernacular names were associated with the majority of records retrieved (54 to 83% of total record-keyword results, Table 3), and many specimens with a use-related keyword also contained a vernacular name. Although our quantification is challenged by the difficulty of recovering diverse transcriptions of diverse local names, it is clear that a wealth of local species concepts and linguistic information is attached to specimens.

Other use categories in Table 3 varied in the relative numbers of records they identified. This likely represents different challenges in the ability of search terms to locate records within different use categories, as well as underlying real differences in what sorts of biocultural information botanists record.

3.4 | Some geographic and temporal trends in biocultural specimens

The countries with the most well-represented biocultural collections are typically those with a long-standing institutional focus. For instance, the results in Table 2 from MO show significant numbers of specimens from active or historical programs in Bolivia (ca. 1,500 specimens), Madagascar (ca. 1,500), Peru (3,000), and the US (2,000). The NY collections show strong representation from Brazil (4,462 specimens), Myanmar (957), the US (5,617), and Vanuatu (3,340). Interestingly, Tanzania also contributed 1,400 biocultural specimens, despite not being a primary geographic focus of NY's collecting efforts. Spot-checking of these specimens revealed that they were primarily obtained through exchanges with the University of California, reinforcing the importance and challenges of institutional exchanges in making knowledge available.

Temporal analysis of the MEXU Maximino project database showed that before the 1980s, less than 0.5% of the specimens documented some use, while in later decades, this proportion rose to 2.0%, suggesting a growing emphasis on ethnobotanical research. Individual collector behavior — driven by mentorship lineages and institutional cultures — was an important driver of the type of information that was recorded on specimen labels.

4 | DISCUSSION

4.1 | The hidden potential of ethnobotanical data in herbaria

Our case study confirmed considerable variation in how ethnobotanical data is recorded across different herbaria. Nevertheless, keyword search terms identified potential ethnobotanical data in about 1.6% of the total specimens analyzed. While this percentage is small, it suggests that the more than 20 million specimens stewarded by these herbaria include hundreds of thousands of specimens containing ethnobotanical data. Extrapolating this projection to the estimated 350 to 400 million specimens housed in herbaria worldwide suggests that over one million may contain associated ethnobotanical data, supporting our central hypothesis that these collections represent an underutilized source of biocultural information.

This potential exists despite ethnobotanical data being, for many herbaria, an unanticipated application of botanical specimens. Although botanical institutions have historical connections to the study of medicinal plants, most contemporary herbaria prioritize taxonomic and systematic objectives. This can mean that some specimens (e.g. fragmentary or sterile specimens) that are important vouchers for ethnobotany may not be accepted into herbaria, or may not be retrievable within curation structures built for botanical records. These challenges are not limited to ethnobotany but also affect plant collections containing vouchers for ecological studies (Meineke & Daru, 2021; Panchen et al., 2019). Neglecting ethnobotanical voucher specimens limits opportunities for interdisciplinary research and disrupts critical connections to cultural knowledge. Incorporating these often-overlooked cultural dimensions into digital herbarium collections, along with careful assessment of potential data and specimen biases, can help bridge this gap (e.g., Panchen et al., 2019).

4.2 | Addressing challenges through compatibility

An important barrier to effectively utilizing ethnobotanical data is the lack of inter-institutional compatibility in herbarium classification systems. Developing institutional practices that are explicitly designed to align with those of other institutions may offer one path forward. In parallel, some new institutional practices aim to integrate Indigenous and local classifications of plants into their herbarium databases. For example, in Brazil, researchers from the Rio de Janeiro Botanical Garden and the National Indigenous Peoples Foundation (FUNAI) are developing a controlled vocabulary for ethnobotany that links established academic analytical categories (e.g., those proposed by Cook, 1995) with culturally-specific terms drawn from key publications documenting plant use in traditional communities (e.g., Araújo, 1961; Cavalcante & Frikel, 1973; Dias & Laureano, 2009). This ongoing initiative also involves evaluating lexical management tools to develop more comprehensive and culturally responsive knowledge repositories. Particular emphasis is placed on preserving Indigenous phytonyms, which are recognized as important carriers of traditional ecological knowledge and plant use practices.

We also observed significant variability in the quality of biocultural information stored in herbaria. Errors can result from researchers documenting data outside their areas of expertise or formal training. A particular challenge lies with vernacular names, which are often collected without proper linguistic training, leading to inaccuracies, misunderstandings, or misrepresentations. This issue is particularly problematic when names are phonetically transcribed, when botanists or local experts misinterpret the intent of the questions being asked, or when a language is dormant. The challenges related to linguistic accuracy in older collections, as well as the difficulties encountered at various stages of transcribing vernacular names, demonstrate the potential in revisiting historical plant collections. This process should be paired with best practices for using linguistic data repositories and collaborations with speaker communities, language keepers, and linguists (Hart & Harrison, 2024; Molino et al., 2022).

Bycatch (false positives that do not represent ethnobotanical data) is a major challenge that will require ongoing refinement of search methodologies to ensure accuracy and ethnobotanical relevance. Refining searches through post-processing with regular expressions to remove some high-frequency bycatch words can increase the accuracy rate, yet a manual review of retrieved records will remain necessary. However, the most time-consuming aspect of previous digitization efforts has been photographing and transcribing specimen information. Manual record reviews are generally quick for individual specimens but can be optimized by prioritizing those that meet multiple keyword search criteria. Rapidly advancing automated systems (large language models and other 'AI') are likely to facilitate increasingly sophisticated transcription, categorization, and pattern analysis of all types of digital herbarium specimen data.

4.3 | Ethical implications and the path forward

Digitization of ethnobotanical data raises critical ethical questions around the recognition and sovereignty of source individuals and communities who have contributed these data (Levis et al., 2024). Many earlier collections were created without the explicit or documented consent or acknowledgment of these communities. Moreover, Open Access and AI initiatives and granting requirements that promote data accessibility risk exposing culturally sensitive knowledge that may be considered sacred or proprietary by IPLCs (Walter et al., 2021). Digitization can provide some form of indirect protection for the biocultural information described in databases via international standards on copyrights. However, this protection does not cover the totality of associated practices, which might be appropriated without the consent of IPLCs (WIPO, 2023). Additionally, herbaria regularly exchange duplicate specimens with each other as a way to broaden access and protect information against loss by housing duplicates in multiple locations. However, associated rights, restrictions, or consent documentation are not always shared as part of these exchanges, particularly in earlier periods. As a result, institutions that receive duplicates must also be provided with, and follow, restriction information.

To address these concerns, there is a pressing need to respect data sovereignty and adopt data governance principles. These may include controlled-access systems for culturally sensitive information and mechanisms for the rematriation/repatriation of data (Leonard, 2024; McAlvay et al., 2021). Future efforts should prioritize digitization co-curation with source communities, along with initiatives such as the Biocultural and Traditional Knowledge Labels, which provide an effective way for communities to assert ownership over their knowledge (Anderson & Hudson, 2020). Moreover, integrating IPLCs' perspectives into the digitization process can support alignment with the Nagoya Protocol's principles of benefit-sharing and fair use. Some projects aspire to actively involve IPLCs in the determination of data architecture (Anderson & Hudson, 2020; Beelbin et al., 2021). Pilot projects focusing at first on specific plant families or localities are currently underway at several institutions (e.g., NY is piloting labels through Local Contexts, https://localcontexts.org/).

4.4 | Geographic limitations

We acknowledge the importance of including the perspectives and data of botanical institutions and scholars from Asia and Africa, two regions with exceptionally rich traditions of botanical knowledge and extensive herbarium collections. Expanding these results and mapping best practices in collaborative groups that include institutions from these regions is a critical next step that we are actively planning for in future work.

5 | CONCLUSIONS

5.1 | Toward best practices and their implementation

Our findings highlight the urgent need to develop and implement best practices for empowering ethnobotanical data in herbaria. A clear next step is the creation of guidelines for recording, processing, analyzing, accessing, and sharing such data—both within and among herbaria, source communities, and the broader public— while promoting a respectful, ethical, and inclusive approach to the digitization of biocultural collections. Key elements of this envisioned framework include:

- Communication, coordination, and mutual support among interested herbaria across the world, with active participation of source communities as partners in the co-production of knowledge;
- Enhanced digitization protocols that prioritize comprehensive ethnobotanical data capture ("no data left behind");
- Context-sensitive metadata standards designed to reflect cultural nuances, community preferences, and principles of local ownership, allowing, where applicable, a process for community curation;
- Ethical guidelines in support of IPLCs' worldviews, data access protocols, intellectual property rights, international agreements, and mechanisms for the return (rematriation and repatriation) of data to IPLCs.

Expanded collaborative work towards these goals is increasingly timely in this period of mass specimen digitization and is urgently needed to strengthen the role of herbaria in biodiversity conservation and cultural resilience.

AUTHOR CONTRIBUTIONS

All authors planned and designed the research. RH, VF-K, MP, MS, LB-R, MM-R, DGF-C, ED-LAES, CM-S, MN, CR, AT, AP, IV performed the research or collected, interpreted, and analyzed the data. RH and IV coordinated writing. All authors contributed to the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge and thank the many individuals and communities from whom and where this biocultural information originated. We also thank the field collectors, curators, collections, and digitization staff at the included herbaria for making these specimens and data available, and Laura Green for analyzing data from Kew. Recent digitization and curation of biocultural specimens at MO was made possible by NEH award ZPA-283531-22 and at NY was made possible by NSF award 2140478.

CONFLICT OF INTEREST STATEMENT

Authors declare there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Tables 1, 2, and 3, and at the public domain resources listed in Table 2.

ORCID

Robbie Hart https://orcid.org/0000-0002-8803-8059
Viviane Fonseca-Kruel https://orcid.org/0000-0001-8647-9734
Eduardo Dalcin https://orcid.org/0000-0002-4661-0272
Estevão da Silva Luís Alexandre https://orcid.org/0000-0002-7464-922X
Matthew Pace https://orcid.org/0000-0002-2333-1924
Michaela Schmull https://orcid.org/0009-0007-8678-7033
Leonardo Beltran-Rodríguez https://orcid.org/0000-0001-6291-1000
Miguel Murguía-Romero https://orcid.org/0000-0002-2532-7398
Diana G. Flores-Camargo https://orcid.org/0009-0009-6979-5116
Cristina Mapes-Sánchez https://orcid.org/0009-0002-5194-5277
Mark Nesbitt https://orcid.org/0000-0002-9588-4296
Carolina Romero https://orcid.org/0000-0001-6194-744X
Andrew Townesmith https://orcid.org/0000-0001-7587-2776
Alex McAlvay https://orcid.org/0000-0001-7051-2018

Kenneth R. Otero-Walker https://orcid.org/0000-0001-9404-1505
Michael J. Balick https://orcid.org/0000-0003-3729-1756
Jacob Golan https://orcid.org/0000-0002-9127-1329
Bruce Hoffman https://orcid.org/0000-0003-4847-5163
Kelsey Leonard https://orcid.org/0000-0002-7531-128X
Giulia Mattalia https://orcid.org/0000-0002-1947-7007
Guillaume Odonne https://orcid.org/0000-0003-1643-309X
Aurora Prehn https://orcid.org/0000-0001-5867-9165
Ina Vandebroek https://orcid.org/0000-0002-1663-0045

REFERENCES

Anderson, J., & Hudson, M. (2020). The biocultural labels initiative: Supporting indigenous rights in data derived from genetic resources. *Biodiversity Information Science and Standards*, *4*, e59230. https://doi.org/10.3897/biss.4.59230

Araújo, A. M. (1961). *Medicina Rústica*. Companhia Editora Nacional.

Arrivabene, A., Lasic, L., Blanco, J., Carrière, S. M., Ladio, A., Caillon, S., Porcher, V., & Teixidor-Toneu, I. (2024). Ethnobiology's contributions to sustainability science. *Journal of Ethnobiology*, *44*(3), 207–220. https://doi.org/10.1177/02780771241261221

Bebber, D. P., Carine, M. A., Wood, J. R., Wortley, A. H., Harris, D. J., Prance, G. T., Davidse, G., Paige, J., Pennington, T. D., Robson, N. K., & Scotland, R. W. (2010). Herbaria are a major frontier for species discovery. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 22169–22171. https://doi.org/10.1073/pnas.1011841108

Beelbin, L., Wallis, E., Hobern, D., & Zerger, A. (2021). The atlas of living Australia: History, current state and future directions. *Biodiversity Data Journal*, *9*, e65023. https://doi.org/10.3897/BDJ.9.e65023

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, *19*(43), 1–12. https://doi.org/10.5334/dsj-2020-043

Cavalcante, P. B., & Frikel, P. (1973). *A Famacopéia Tiriyó: estudo étno-botânico*. Museu Paranaense Emílio Goeldi.

Cook, F.E.M. (1995). Economic botany data collection standard. Royal Botanic Gardens, Kew. Accessed May 29, 2025. http://www.tdwg.org/standards/103

Cowell, C., Paton, A., Borrell, J. S., Williams, C., Wilkin, P., Antonelli, A., Baker, W. J., Buggs, R., Fay, M. F., Gargiulo, R., Grace, O. M., Kuhnhäuser, B. G., Woudstra, Y., & Kersey, P. J. (2021). Uses and benefits of digital sequence information from plant genetic resources: Lessons learnt from botanical collections. *Plants, People, Planet*, *4*(1), 33–43. https://doi.org/10.1002/ppp3.10216

Davis, C. C., Sessa, E., Paton, A., Antonelli, A., & Teisher, J. K. (2025). Guidelines for the effective and ethical sampling of herbaria. *Nature Ecology & Evolution*, *9*, 196–203. https://doi.org/10.1038/s41559-024-02544-z

Delves, J., Albán-Castillo, J., Cano, A., Aviles, C. F., Gagnon, E., Gonzáles, P., Knapp, S., León, B., Marcelo-Peña, J. L., Reynal, C., Gonzáles, R. P. R., Rodrígues, E. F. R., Särkinen, T., Martinez, R. V., & Moonlight, P. W. (2024). Small and in-country herbaria are vital for accurate plant threat assessments: A case study from Peru. *Plants, People, Planet*, *6*, 174–185. https://doi.org/10.1002/ppp3.10425

Dias, J. E., & Laureano, L. C. (2009). *Farmacopeia Popular do Cerrado*. Articulação Pacari.

Flores-Camargo F.G. Sánchez-Dirzo M.G. 2022 Maximino: Base de datos de plantas útiles de México. www.maximinom.org

Friedl, J. E. F. (2006). *Mastering regular expressions*. O'Reilly.

Hart, R., & Harrison, K. D. (2024). Introduction: Languages and plants together promote planetary health. *Sociolinguistic Studies*, *18*(3–4), 307–316. https://doi.org/10.3138/ss-18-3.4-0001

Kuhlman, A., & Salick, J. (2014). Database standards for biocultural collections. In J. Salick, K. Konchar, & M. Nesbitt (Eds.), *Curating biocultural collections: A handbook* (pp. 151–168). Kew Publishing with Missouri Botanic Garden.

Leonard, K. (2024). Decolonizing botanical gardens. *Qualitative Research Journal*, *24*(5), 536–554. https://doi.org/10.1108/QRJ-04-2024-0086

Levis, C., Hanazaki, N., Zank, S., Peroni, N., Julião, C. G., Da Silva, M. T., De Alencar Assis, A. L. A., Nakamura, E. M., Soldati, G., Almada, E. D., Odonne, G., & Teixidor-Toneu, I. (2024). Safeguard stewards of biodiversity knowledge. *Science*, *385*(6708), 504–506. https://doi.org/10.1126/science.adp1749

McAlvay, A. C., Armstrong, C. G., Baker, J., Black Elk, L., Bosco, S., Hanazaki, N., Joseph, L., Martínez-Cruz, T. E., Nesbitt, M., Palmer, M. A., Priprá de Almeida, W. C., Anderson, J., Asfaw, Z., Borokini, I. T., Cano-Contreras, E. J., Hoyte, S., Hudson, M., Ladio, A. H., Odonne, G., … Vandebroek, I. (2021). Ethnobiology phase VI: Decolonizing institutions, projects, and scholarship. *Journal of Ethnobiology*, *41*, 170–191. https://doi.org/10.2993/0278-0771-41.2.170

Meineke, E. K., & Daru, B. H. (2021). Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution*, *36*(12), 1071–1082. https://doi.org/10.1016/j.tree.2021.08.003

Molino, J.-F., Sabatier, D., Grenand, P., Engel, J., Frame, D., Delprete, P. G., Fleury, M., Odonne, G., Davy, D., Lucas, E. J., & Martin, C. A. (2022). An annotated checklist of the tree species of French Guiana, including vernacular nomenclature. *Adansonia*, *44*(26), 345–903. https://doi.org/10.5252/adansonia2022v44a26

Murguía-Romero, M., Serrano-Estrada, B., Salazar, G. A., Sánchez-González, G. E., Palacios, U. M. S., Gernandt, D. S., Magallón, S., & Sánchez-Cordero, V. (2024). The IBdata web system for biological collections: Design focused on usability. *Biodiversity Informatics*, *18*, 1–12. https://doi.org/10.17161/bi.v18i.20516

Nesbitt, M. (2014). Use of herbarium specimens in ethnobotany. In J. Salick, K. Konchar, & M. Nesbitt (Eds.), *Curating biocultural collections: A handbook* (pp. 313–328). Kew Publishing with Missouri Botanic Garden.

Odonne, G., Tareau, M.-A., & Van Andel, T. (2021). Geopolitics of bitterness: Deciphering the history and cultural biogeography of *Quassia Amara* L. *Journal of Ethnopharmacology*, *267*, 113546. https://doi.org/10.1016/j.jep.2020.113546

Panchen, Z. A., Doubt, J., Kharouba, H. M., & Johnston, M. O. (2019). Patterns and biases in an Arctic herbarium specimen collection: Implications for phenological research. *Applications in Plant Sciences*, *7*(3), e01229. https://doi.org/10.1002/aps3.1229

Primack, D., Imbres, C., Primack, R. B., Miller-Rushing, A. J., & Del Tredici P. (2004). Herbarium specimens demonstrate earlier flowering times in response to warming Boston. *American Journal of Botany*, *91*, 1260–1264.

Salick, J., Konchar, K., & Nesbitt, M. (Eds.). (2014). *Curating biocultural collections: A handbook*. Kew Publishing with Missouri Botanic Garden.

Secretariat of the Convention on Biological Diversity. (1992). The Convention on Biological Diversity. Accessed May 29 2025 https://www.cbd.int/youth/0003.shtml

Silva, L. A. E., Fraga, C. N., Almeida, T. M. H., Gonzalez, M., Lima, R. O., Rocha, M. S., Bellon, E., Ribeiro, R. S., Oliveira, F. A., Clemente, L. S., Magdalena, U. R., Medeiros, E. V. S., & Forzza, R. C. (2017). Jabot-Sistema de Gerenciamento de Coleções Botânicas: A experiência de uma década de desenvolvimento e avanços. *Rodriguésia*, *68*, 391–410. https://doi.org/10.1590/2175-7860201768208

Thiers, B. M. (2020). *Herbarium: The quest to preserve and classify the world's plants*. Timber Press. Accessed May 29 2025 https://scholar.google.com/scholar?cluster=12941843400798571045&hl=en&oi=scholar

Thiers, B. M. . 2025 The World's Herbaria 2024: A Summary Report Based on Data from Index Herbariorum. New York, New York: New York

Botanic Garden, January 2025. Accessed May 29, 2025 https://sweetgum.nybg.org/science/wpcontent/uploads/2025/01/The_World_Herbaria_2024-.pdf

Vandebroek, I., Picking, D., Aiken, S., Lewis, P. A., Oberli, A., Mitchell, S., & Boom, B. (2018). A review of Coralilla (*Antigonon leptopus*): An invasive and popular urban bush medicine in Jamaica. *Economic Botany*, *72*(2), 229–245. https://doi.org/10.1007/s12231-018-9415-5

von Reis, S. (Ed.). (1973). *Drugs and foods from little-known plants: Notes in Harvard University herbaria*. Harvard University Press.

von Reis, S., & Lipp, F. J. (1982). *New plant sources for drugs and foods from the New York botanical garden herbarium*. Harvard University Press.

Walter, M., Lovett, R., Maher, B., Williamson, B., Prehn, J., Bodkin-Andrews, G., & Lee, V. (2021). Indigenous data sovereignty in the era of big data and open data. *Australian Journal of Social Issues*, *56*(2), 143–156. https://doi.org/10.1002/ajs4.141

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Cruzes, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … González-Beltran, A. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

World Intellectual Property Organization. (2023). Intergovernmental committee on intellectual property and genetic resources, traditional knowledge and folklore, the protection of traditional knowledge: Updated draft gap analysis. WIPO/GRTKF/IC/47/8. Accessed May 29, 2025 https://www.wipo.int/edocs/mdocs/tk/en/wipo_grtkf_ic_47/wipo_grtkf_ic_47_8.pdf

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hart, R., Fonseca-Kruel, V., Dalcin, E., da Silva Luís Alexandre, E., Pace, M., Schmull, M., Beltran-Rodríguez, L., Murguía-Romero, M., Flores-Camargo, D. G., Mapes-Sánchez, C., Nesbitt, M., Romero, C., Townesmith, A., Salick, J., McAlvay, A., Otero-Walker, K. R., Balick, M. J., Golan, J., Hoffman, B., … Vandebroek, I. (2025). Repositories of biocultural diversity: Toward best practices for empowering ethnobotany in digital herbaria. *Plants, People, Planet*, 1–9. https://doi.org/10.1002/ppp3.70052