ARPHA Preprints

PENSOFT

PREPRINT

Author-formatted, not peer-reviewed document posted on 04/05/2026

DOI: https://doi.org/10.3897/arphapreprints.e197961

Digitising historical plant collections in Zimbabwe: Implications for biodiversity data mobilisation and global plant knowledge

Kudakwashe Mutasa

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 04/05/2026. DOI: https://doi.org/10.3897/arphapreprints.e197961

Digitising Historical Plant Collections in Zimbabwe: Implications for Biodiversity Data Mobilisation and Global Plant Knowledge

Introduction

Herbaria provide the empirical foundation for botanical science, offering verifiable records of species identity, distribution and ecological context through time (Funk 2003). These collections are central to plant systematics because they preserve the physical evidence upon which taxonomic concepts, nomenclature and species descriptions are based (Bebber et al. 2010). As a result, herbarium specimens continue to play a critical role in documenting plant diversity and supporting taxonomic revisions, floristic studies and conservation assessments. Increasingly, they are also used to address broader scientific questions related to global environmental change, including shifts in species distributions, phenological responses to climate change and patterns of plant invasions (Soltis 2017, Heberling et al. 2019).

Globally, approximately 396 million preserved plant specimens are preserved in the world's 3,000 active herbaria (Roma-Marzio et al. 2023, De Smedt et al. 2024) and provide reference material for species identification, distribution, ecology, typification, morphological comparisons and phylogenetic studies. Recent advances in biodiversity informatics have significantly enhanced the research value of herbarium collections. Digitisation, defined as the conversion of physical specimens and their associated metadata into digital formats, has enabled large-scale accessibility and integration of herbarium data across institutions and countries (Nelson and Ellis 2019). Through digitisation, specimen images and label information can be shared through global biodiversity data infrastructures, facilitating new forms of research that rely on aggregated datasets spanning multiple collections.

Despite these advances, the global digitisation of herbarium collections remains uneven. While many institutions in Europe and North America have made substantial progress, collections in biodiversity-rich regions of the Global South remain only partially digitised due to financial, technical and infrastructural constraints (Nelson and Ellis 2019). This imbalance limits the representation of biodiversity from these regions in global data systems.

Zimbabwe provides an illustrative case study of digitisation in resource-constrained environments. The National Herbarium and Botanic Garden (SRGH) houses one of the largest herbarium collections in southern Africa, yet much of this material has been accessible primarily through physical consultation. Over the past two decades, multiple initiatives have sought to digitise and mobilise portions of these collections.

This paper examines the digitisation of historical plant collections held at the National Herbarium and Botanic Garden of Zimbabwe (SRGH) and evaluates the outcomes of these initiatives in improving accessibility and use of herbarium data. It further discusses the institutional and structural challenges encountered during digitisation and considers the broader implications of these experiences for strengthening global plant knowledge mobilisation.

History of the Herbarium Collection

SRGH houses over 500,000 vascular plant specimens dating to the early nineteenth century. During the colonial period, the herbarium served as the central repository for the collections made in the Rhodesia-Nyasaland Federation, which consisted of Malawi, Zambia and Zimbabwe; hence it has a vast collection of historical collections from the southern African region, were it is ranked

as the third largest herbarium. The herbarium, which is the national repository of plant data, contains specimens representing a wide diversity of Zimbabwean flora, including vascular plants from different ecological regions of the country. These specimens provide historical records of plant distribution and serve as important reference material for taxonomic studies.

Much of this material was collected prior to the widespread use of geospatial technologies, resulting in handwritten labels with locality descriptions that vary in precision. While scientifically invaluable, such analogue collections face risks of physical degradation and remain poorly integrated into modern, data-driven research and policy frameworks without digitisation.

Digitisation of Herbarium collection

Digitisation of Zimbabwe's plant collections has been driven by three interlinked imperatives. First, there is increasing national demand for reliable biodiversity data to support conservation planning, environmental impact assessments and reporting obligations under multilateral environmental agreements such as the Convention on Biological Diversity. Second, global biodiversity platforms, including the Global Biodiversity Information Facility and initiatives such as the World Flora Online, have created both incentives and standards for sharing specimen data internationally. Third, digitisation serves as a preservation strategy by reducing handling of fragile specimens while safeguarding associated information against loss. In this context, digitisation represents not merely a technical exercise but a strategic investment in national botanical infrastructure.

Efforts to digitise the collection began in the early 2000s with support from the Southern African Botanical Diversity Network (SABONET) project (Willis and Huntley 2001). This initiative contributed to capacity development among herbarium staff in digitisation skills and supported the development of key outputs, including a comprehensive checklist of Zimbabwean plants (Mapaura and Timberlake 2004) and national Red Data Lists (Golding, 2002). Following the conclusion of SABONET in 2005, digitisation efforts continued under the African Plants Initiative (API) (Smith et al. 2011), which brought together a consortium of African herbaria to digitise type specimens (specimens representing the original material used in species descriptions). This initiative has since expanded to a global project, Global Plants Initiative.

These early digitisation efforts utilised the Pretoria Computerised Information System (PRECIS), a database developed by the South African National Herbarium. However, continued use of this system became challenging due to technological advancements that rendered both the software and its supporting operating systems obsolete. In 2023, the herbarium secured grant support to facilitate migration from PRECIS to the Botanical Research and Herbarium Management System (BRAHMS) through the project "Strengthening institutional capacity of the National Herbarium as an Information Hub for Plant Genetic Resources in Zimbabwe," supported by the Seed Innovation Fund of the Convention on Biological Diversity. A key component of this project involved capacity development in database management for herbarium personnel.

Additional initiatives have also contributed to the digitisation of specialised plant groups. For example, funding from the Max Planck Institute supported the digitisation of crop wild relatives,

resulting in approximately 2,000 specimens being digitised. Collectively, these initiatives have progressively strengthened the digitisation and data mobilisation capacity of the herbarium.

Digitisation Approaches and Workflows

Digitisation efforts in Zimbabwe have focused on specimen imaging and data capture using internationally recognised standards. High-resolution images are generated using both flatbed scanners and camera-based systems (Figure 1), with prioritisation given to type specimens, which are a historically significant collection and taxa of conservation concern

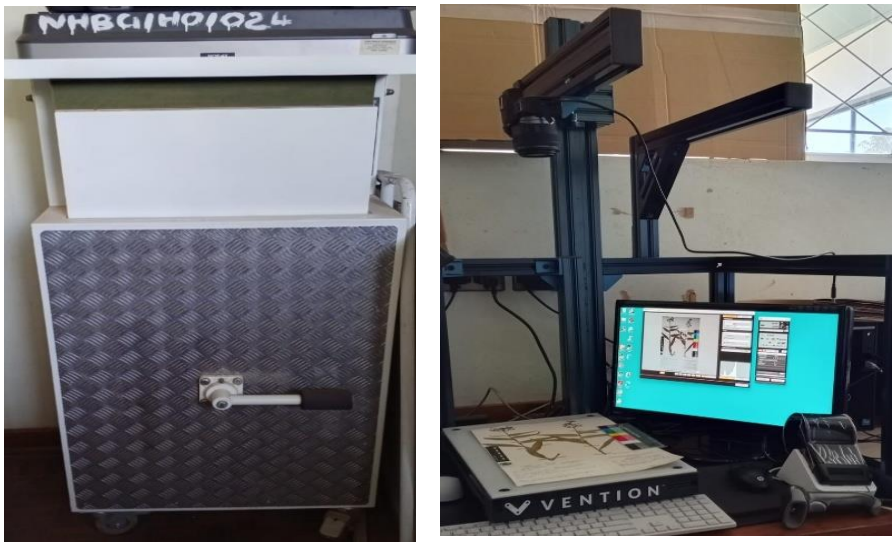

*Figure 1. Flatbed scanner and camera-based digitisation workstations employed at the **National Herbarium and Botanic Garden** (SRGH) to support high-resolution specimen imaging during the digitisation workflow.*

The digitisation workflow (Figure 2) follows a structured sequence designed to ensure data quality and long-term usability. Specimens are first selected and prepared, including verification of taxonomic names and physical condition. Each specimen is then imaged using either a flatbed scanner or a camera-based workstation to produce high-resolution digital images. Label information is subsequently transcribed into a database, capturing key fields such as collector name, collection number, locality, date and taxonomic identification, following the Darwin Core standard (Wieczorek et al. 2012). Geographic information is georeferenced with the assistance of online mapping tools, historical maps, gazetteers and expert knowledge where possible. Records are subject to quality control to verify completeness and accuracy. Finally, validated records are integrated into the herbarium database and prepared for sharing through international biodiversity data platforms, enhancing accessibility and reuse for research and conservation.

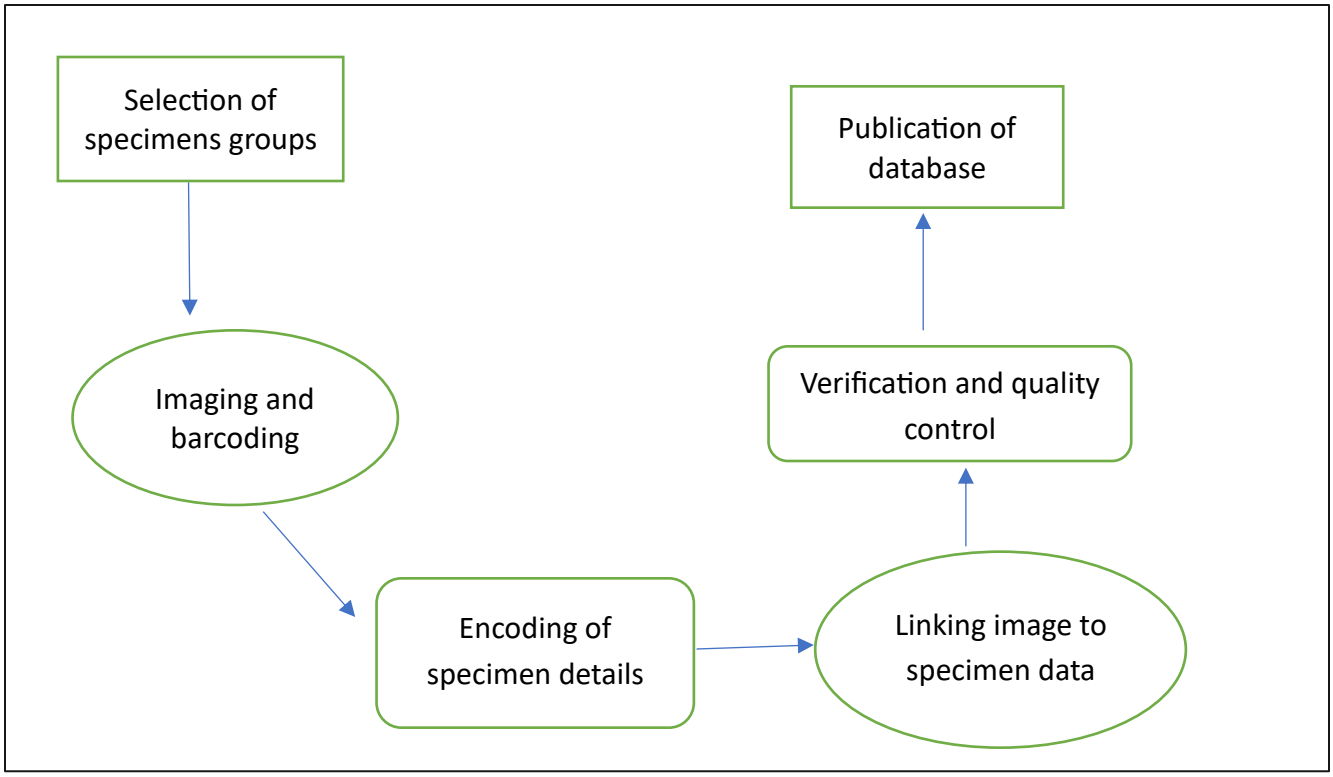

Figure 2. Workflow for digitization of herbarium specimens

Outcomes and Scientific Value

Digitisation has substantially increased the accessibility and scientific use of Zimbabwe's plant collections, while revealing important limitations in downstream application. To date, approximately 20% of the herbarium collection, representing 109 218 specimen records, has been transcribed into digital format. The digitised dataset is taxonomically uneven, with the families Poaceae and Fabaceae accounting for the largest proportions of digitised material, contributing 17.1% and 12.5% of the records respectively. Other well-represented families include Asteraceae, Cyperaceae and Rubiaceae (Table 1). This pattern reflects prioritisation of ecologically dominant and economically important groups, as well as the availability of curated material suitable for digitisation.

Table 1. Taxonomic distribution of the ten most represented plant families in the digitised herbarium collection from Zimbabwe.

Family	Number of Specimens	Proportion (%)
Poaceae	18735	17.15%
Fabaceae	13659	12.51%
Asteraceae	6663	6.10%
Cyperaceae	4249	3.89%
Rubiaceae	4067	3.72%
Orchidaceae	3416	3.13%
Euphorbiaceae	3364	3.08%
Lamiaceae	2543	2.33%
Acanthaceae	2451	2.24%
Apocynaceae	2385	2.18%

Type specimens, which are of particular importance in plant systematics and nomenclatural research, constitute a relatively small proportion of the digitised collection, accounting for 1,673 specimens (1.53% of the total digitised records). Among the digitised type material, isotypes form the largest component, followed by type specimens and holotypes, while paratypes, syntypes and other categories such as isosyntypes, co-types and neotypes are represented in much smaller proportions (Figure 3).

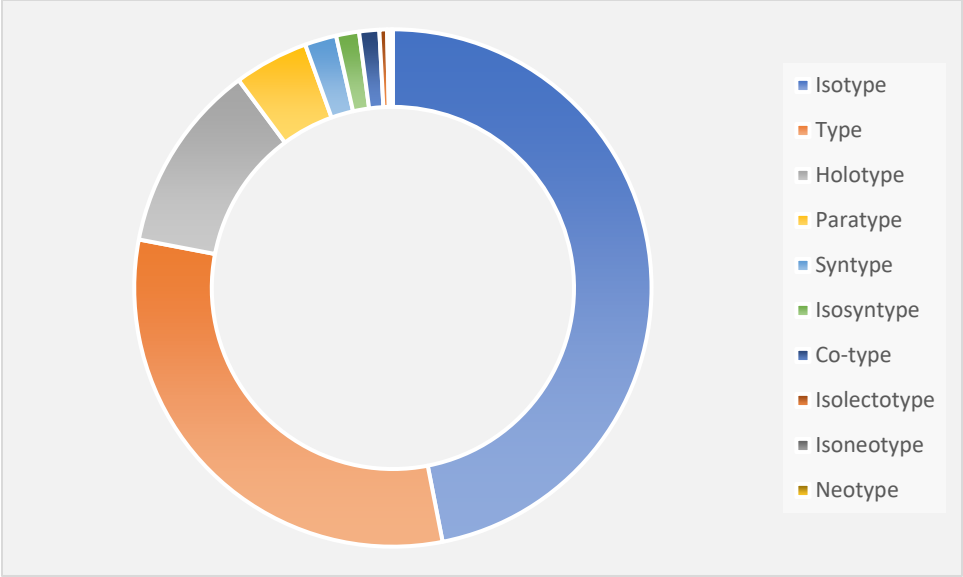

Figure 3. Distribution of digitised type specimens by category in the Zimbabwe herbarium.

Despite progress in digitisation, significant gaps remain in the mobilisation and publishing of Zimbabwean plant specimen data. Although digitised records are accessible through international biodiversity platforms such as the Global Biodiversity Information Facility and the Global Plants Initiative, analysis of datasets available through GBIF indicates that approximately 138,671 occurrence records relating to Zimbabwean plants have been published for global access (Figure 4), with the majority originating from institutions outside Zimbabwe. Institutions in the United Kingdom account for 45% (62,688 records) of the dataset, followed by the United States with 18% (25,374 records). Within the region, South Africa contributes 9.9% (13,765 records), whereas

Zimbabwe contributes only 5.3% (7,367 records). This imbalance highlights a gap in the digitisation and data publishing process, reflecting both the historical dispersal of specimens to foreign institutions and the limited national capacity for large-scale data mobilisation.

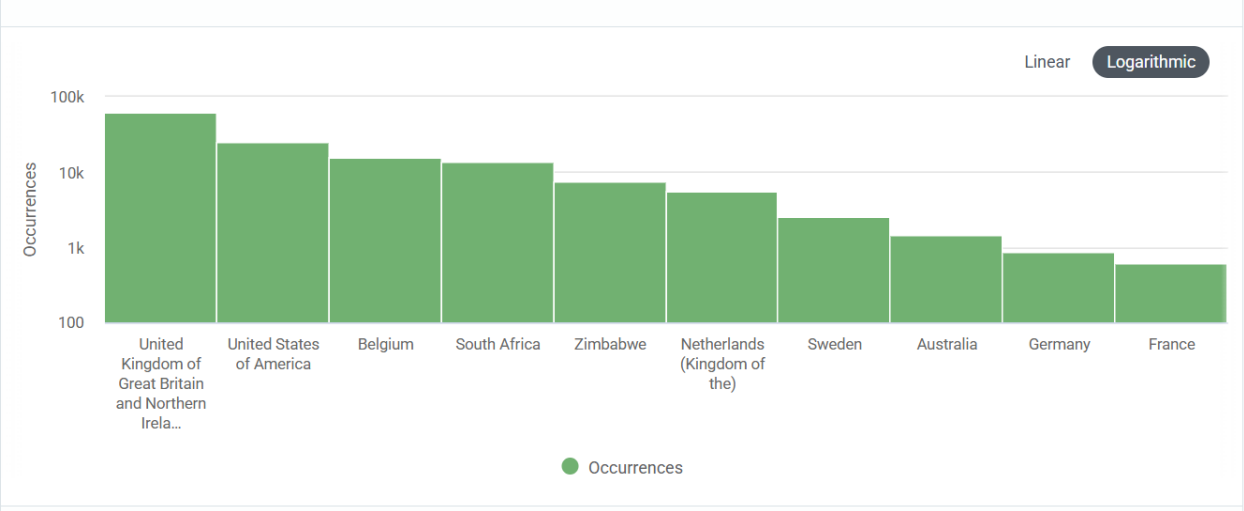

Figure 4. Distribution of specimen records on GBIF platform by publishing country. (Source: www.gbif.org)

In addition to occurrence records available through GBIF, digitised images and metadata for approximately 3,000 type specimens from Zimbabwe are accessible through the Global Plants Initiative, further enhancing global access to taxonomically important material held in the collection. The availability of digitised herbarium data has significantly enhanced opportunities for biodiversity research, including species distribution studies, conservation assessments and ecological analyses. However, the dominance of foreign institutions in publishing occurrence records relating to Zimbabwean plants highlights broader issues concerning historical specimen dispersal and the ownership of biodiversity data. Strengthening national digitisation capacity will therefore be critical for improving the visibility and stewardship of plant biodiversity information originating from Zimbabwe.

A search conducted using the Google Search engine in the fields of ethnobotany, biodiversity conservation and policy in Zimbabwe indicates that herbarium-derived data are used in a range of applied research areas, although their overall utilisation remains limited. The search results show that herbarium data has the potential to contribute to studies on ethnobotany, plant genetic resources, invasive alien species, floristic inventories and baseline biodiversity assessments (Maroyi 2012, 2017, Alfred 2022, Mutasa et al. 2025). These findings highlight the growing, yet still underutilised, role of herbarium data in supporting biodiversity research and policy development. Digitisation has also enhanced the potential for integrating Zimbabwean plant records into large-scale global analyses. Increasingly, specimen occurrence data are used in species distribution modelling, climate change impact assessments and conservation prioritisation exercises. Inclusion of Zimbabwean specimen records within global datasets improves the representation of southern African flora in such analyses and helps address geographic biases that have historically characterised biodiversity data availability.

Despite these advances, the analysis also reveals important limitations in the downstream use of digitised data within national policy and conservation planning processes. Explicit examples of

digitised herbarium data informing national biodiversity strategies, protected area planning, extinction risk assessments or environmental decision-making remain relatively limited. This research–policy gap reflects broader challenges in translating scientific data into policy-relevant outputs and highlights the need to strengthen institutional linkages between research institutions and environmental governance structures.

Beyond data mobilisation, digitisation initiatives have generated important human capital benefits within the herbarium. Staff members have received training in specimen imaging, biodiversity data standards, georeferencing techniques and database management. These capacity gains are particularly significant for institutions in the Global South, where specialised expertise in biodiversity informatics remains limited. Developing these skills is essential for ensuring that herbaria continue to function as active research infrastructures rather than static repositories of historical collections.

Persistent Challenges

While significant progress has been achieved in digitising Zimbabwe's herbarium collections, several structural constraints continue to limit the scale and sustainability of digitisation efforts. These challenges reflect broader patterns experienced by many natural history institutions in the Global South.

One of the most significant constraints is the reliance on short-term, project-based funding. Many digitisation initiatives have been supported through externally funded projects, which often operate within limited timeframes. While such projects can produce important outputs, including specimen imaging, database development and staff training, they rarely provide the long-term financial support required for sustained data curation, system maintenance and continued digitisation of large collections.

Infrastructure limitations also present ongoing challenges. Digitisation workflows require a reliable electricity supply, adequate computing infrastructure and sufficient data storage capacity. In contexts where infrastructure is inconsistent, these requirements can slow digitisation processes and complicate long-term data management.

Human resource constraints represent another significant barrier. Although training initiatives have improved institutional capacity, specialised expertise remains limited in areas such as georeferencing, database administration and taxonomic research for certain plant groups. In particular, shortages of specialists in groups such as fungi and cryptogams constrain the ability to comprehensively digitise and verify collections representing these taxa.

In addition, many historical specimens contain locality information recorded as descriptive place names rather than precise geographic coordinates. Georeferencing these records requires substantial time and expert interpretation using historical maps, gazetteers and collector knowledge. In some cases, locality descriptions remain too ambiguous to assign reliable coordinates, limiting the usefulness of these records for spatial analyses.

Another emerging challenge concerns data governance and intellectual property. As digitised specimen data are increasingly shared through global biodiversity platforms, questions arise regarding data ownership, attribution and equitable use. Countries hosting biodiverse ecosystems often contribute large amounts of primary biodiversity data but may lack the analytical capacity or

institutional frameworks needed to fully utilise these datasets. Strengthening national data publishing capacity, therefore becomes important not only for improving data accessibility but also for ensuring appropriate recognition and stewardship of biodiversity information.

Addressing these challenges will require sustained institutional investment, long-term funding strategies and stronger integration of herbarium digitisation into national biodiversity information systems. Without such measures, progress achieved through individual projects may be difficult to maintain over the long term.

Lessons for Global Plant Knowledge

Zimbabwe's experience in digitising historical plant collections provides important insights into the evolving role of herbaria within global biodiversity knowledge systems. As biodiversity research increasingly relies on large-scale digital datasets, the mobilisation of herbarium records from biodiversity-rich regions is becoming essential for improving the geographic and taxonomic coverage of global plant data infrastructures. Digitised specimens contribute to diverse areas of research, including species distribution modelling, conservation assessments, phylogeographic studies and analyses of ecological change over time. Platforms such as the Global Biodiversity Information Facility and the World Flora Online rely heavily on specimen-based data to support global biodiversity assessments and taxonomic synthesis.

However, Zimbabwe's experience also highlights structural inequalities in the global distribution and mobilisation of biodiversity data. Many plant specimens collected from the Global South during earlier periods of botanical exploration are currently housed in herbaria located in Europe and North America. As a result, a substantial proportion of digitised occurrence data for Zimbabwean plants is currently published through institutions outside the country. While these collections remain scientifically valuable, this pattern reflects broader historical imbalances in the control, accessibility and publication of biodiversity information.

Increasing national participation in biodiversity data publishing is therefore critical for ensuring equitable representation within global knowledge systems. Strengthening the capacity of institutions such as the National Herbarium and Botanic Garden of Zimbabwe to digitise, curate and publish specimen data can help ensure that countries hosting significant biodiversity resources also play a leading role in generating and managing the associated knowledge. Such efforts can improve national access to biodiversity information for conservation planning, environmental impact assessment and reporting under international frameworks such as the Convention on Biological Diversity.

Zimbabwe's digitisation initiatives also demonstrate the importance of international partnerships in enabling progress under resource constraints. Collaborative initiatives, including the African Plants Initiative and the Global Plants Initiative, have provided technical expertise, infrastructure and training that have significantly accelerated digitisation efforts. These collaborations illustrate how global networks of herbaria can collectively contribute to improving accessibility of plant collections while supporting capacity development in biodiversity-rich regions.

Looking forward, emerging technologies offer additional opportunities to enhance digitisation efficiency and expand data mobilisation. Tools such as optical character recognition for automated label transcription, machine-learning approaches to georeferencing and improved imaging technologies have the potential to significantly accelerate digitisation workflows. However,

realising these benefits will require sustained institutional investment in infrastructure, training and long-term data stewardship.

Ultimately, Zimbabwe's experience underscores that digitising historical plant collections is not only a technical exercise but also a strategic component of strengthening global plant knowledge. Ensuring that biodiversity-rich countries can effectively digitise, publish and utilise their collections will be essential for building more inclusive and representative biodiversity knowledge systems capable of supporting conservation and sustainable use of plant diversity worldwide.

Conclusion

Digitisation of herbarium collections is increasingly recognised as essential for advancing botanical research, conservation planning and global biodiversity assessments. Zimbabwe experiences how historically analogue plant collections can be transformed into accessible biodiversity data resources through strategic initiatives and the use of open-source data mobilisation facilities. Although significant progress has been made, only a portion of the national collection has been digitised, and challenges related to funding, infrastructure and technical capacity remain. Furthermore, the fact that a large proportion of the country's herbarium specimen data is currently published by institutions outside the country, strengthening national digitisation and data publishing efforts, is critical for improving accessibility, stewardship and ownership of plant biodiversity information. Zimbabwe's experience highlights the importance of sustained investment, partnerships and capacity development in ensuring that biodiversity-rich countries can fully participate in and contribute to global plant knowledge systems.

References

Alfred M (2022) Alien flora of Zimbabwe: Data derived from herbarium specimens. Data in Brief 42: 108186. https://doi.org/10.1016/j.dib.2022.108186

Bebber DP, Carine MA, Wood JRI, Wortley AH, Harris DJ, Prance GT, Davidse G, Paige J, Pennington TD, Robson NKB, Scotland RW (2010) Herbaria are a major frontier for species discovery. Proceedings of the National Academy of Sciences 107: 22169–22171. https://doi.org/10.1073/pnas.1011841108

De Smedt S, Bogaerts A, De Meeter N, Dillen M, Engledow H, Van Wambeke P, Leliaert F, Groom Q (2024) Ten lessons learned from the mass digitisation of a herbarium collection. PhytoKeys 244: 23–37. https://doi.org/10.3897/phytokeys.244.120112

Funk V (2003) The Importance of Herbaria. Plant Science Bulletin 49: 94–95.

Heberling JM, Prather LA, Tonsor SJ (2019) The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis. BioScience 69: 812–822. https://doi.org/10.1093/biosci/biz094

Mapaura A, Timberlake J (Eds) (2004) A checklist of Zimbabwean vascular plants. SABONET, Pretoria and Harare. Southern African Botanical Diversity Network

Maroyi A (2012) The casual, naturalised and invasive alien flora of Zimbabwe based on herbarium and literature records. Koedoe 54: 6 pages. https://doi.org/10.4102/koedoe.v54i1.1054

Maroyi A (2017) Data on introduced plants in Zimbabwe: Floristic changes and patterns of collection based on historical herbarium records. Data in Brief 15: 348–369. https://doi.org/10.1016/j.dib.2017.09.046

Mutasa K, Chipfunde O, Chapano C, Chatukuta P (2025) A national crop wild relative checklist for Zimbabwe reveals edible crop wild relative diversity of regional and global importance. https://doi.org/10.1101/2025.07.04.663189

Nelson G, Ellis S (2019) The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B: Biological Sciences 374: 20170391. https://doi.org/10.1098/rstb.2017.0391

Roma-Marzio F, Maccioni S, Dolci D, Astuti G, Magrini N, Pierotti F, Vangelisti R, Amadei L, Peruzzi L (2023) Digitization of the historical Herbarium of Michele Guadagno at Pisa (PI-GUAD). PhytoKeys 234: 107–125. https://doi.org/10.3897/phytokeys.234.109464

Smith GF, Roux JP (Koos), Raven P, Figueiredo E (2011) African Herbaria Support Transformation on the Continent[1]. Annals of the Missouri Botanical Garden 98: 272–276. https://doi.org/10.3417/2010050

Soltis PS (2017) Digitization of herbaria enables novel research. American Journal of Botany 104: 1281–1284. https://doi.org/10.3732/ajb.1700281

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. Sarkar IN (Ed.). PLoS ONE 7: e29715. https://doi.org/10.1371/journal.pone.0029715

Willis CK, Huntley BJ (2001) SABONET: Developing Capacity within Southern Africa's Herbaria and Botanical Gardens. Systematics and Geography of Plants 71: 247. https://doi.org/10.2307/3668671