Corresponding author: Heru Handika, handika.heru@gmail.com

Running header: NAHPU natural history data acquisition and management

NAHPU: An integrated approach to natural history data acquisition and management in the extended specimen era

Heru Handika[1,2*] (https://orcid.org/0000-0002-2834-7175), Andre E. Moncrieff[3] (https://orcid.org/0000-0002-9783-0400), Cody Henderson[4], Alain Dadaian[4], Caleb M. Gordon[2] (https://orcid.org/0000-0003-3781-7965), Roberta Canton[5] (https://orcid.org/0009-0007-9022-3644), Austin S. Chipps[6](https://orcid.org/0000-0002-0674-9042), Nurul Inayah[7], Ada Chornelia[8](https://orcid.org/0000-0002-8541-5947), Marco A. Rego[2] (https://orcid.org/0000-0002-6783-0012), Natalia Pérez-Amaya[3] (https://orcid.org/0000-0002-3388-256X), Lawrence B. Gan[4] (https://orcid.org/0009-0002-9369-4023), Aakanksha T. Jaiswal[4](https://orcid.org/0009-0007-4809-173X), Daeun D. Lee[4], Jia Yu[4], Ashmitha J. Aravind[4], Dante R. Evangelista[4], Samantha L. Rutledge[1] (https://orcid.org/0000-0001-6351-7785), Diego J. Elías[9] (https://orcid.org/0000-0003-4215-0384), Emily M. Troyer[3] (https://orcid.org/0000-0001-7478-2306), Gustavo S. de Miranda[3] (https://orcid.org/0000-0002-2895-9331), Lauren Vonnahme[10] (https://orcid.org/0009-0002-3253-1343), Breanna Shi[4] (https://orcid.org/0009-0002-5332-4626), Jonathan A. Nations[2] (https://orcid.org/0000-0002-1170-7656), Glaucia Del-Rio[2] (https://orcid.org/0000-0001-7212-8474), Gregory Thom[1] (https://orcid.org/0000-0001-6200-0565), Arthur Porto[2] (https://orcid.org/0000-0002-9210-8750), Jacob A. Esselstyn[1] (https://orcid.org/0000-0002-1823-4062)

[1] Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA.
[2] Florida Museum of Natural History, University of Florida, Gainesville, FL, USA.
[3] Sam Noble Oklahoma Museum of Natural History and School of Biological Sciences, University of Oklahoma, Norman, OK, USA.
[4] College of Computing, Georgia Institute of Technology, Atlanta, GA, USA.
[5] Department of Biological Sciences, Clemson University, Clemson, SC, USA.
[6] National Museum of Natural History, Smithsonian Institution, Washington, DC, USA.
[7] Museum Zoologicum Bogoriense, Research Center for Biosystematics and Evolution, National Research and Innovation Agency (BRIN), Cibinong, Indonesia.
[8] School of BioSciences, The University of Melbourne, Melbourne, VIC, Australia.

[9] Negaunee Integrative Research Center, The Field Museum of Natural History, Chicago, IL, USA.
[10] Department of Herpetology, American Museum of Natural History, New York, NY, USA.

*Current address: Florida Museum of Natural History, University of Florida, Gainesville, FL, USA.

Abstract

Natural history museums collect, curate, and distribute the data that underpin biodiversity science. Modern, holistic collection practices have increased the number of specimen parts, associated parasites, and derivative digital assets, challenging researchers to maintain data integrity for a vast set of data types representing the extended specimen. We developed the NAHPU digital field catalog and supporting software to streamline data collection from the field and upload to museum databases. NAHPU is open source, runs on mobile and desktop operating systems, and is distributed through official app stores with one-click installation and automatic updates on Apple devices, Windows, Android, and Linux. It provides stable support for mammalogy and ornithology, and beta support for herpetology, ichthyology, and invertebrate zoology. In addition to standard data-entry fields, it includes customizable, controlled data-entry vocabularies, flexible exports, a GUI template editor, and a growing list of real-time statistics to support reproducible collection practices. We tested NAHPU on field expeditions in remote, tropical field conditions, including 115 days (993 specimens and 3605 specimen parts) on Sumatra without on-site electricity, and as a long-term data management application. Our approach, involving Flutter, Rust, on-device computing, and open-source, interdisciplinary collaboration, provides a framework for low-cost, open-access software development.

Keywords: biodiversity informatics, digital field catalog, Darwin Core, FAIR data, specimen data management, natural history collections, cross-platform software, extended specimen

Introduction

Natural history collections (NHCs) are vital research and educational infrastructures for documenting biodiversity, environmental change, and disease surveillance (Cook et al. 2014; Schindel and Cook 2018; Boldgiv et al. 2025; Paul et al. 2025). Over the past several decades, large-scale digitization and data mobilization efforts, coordinated through initiatives such as the Integrated Digitized Biocollections (iDigBio; Page et al. 2015), the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), and Symbiota (Gries et al. 2014) have transformed NHCs into integrated systems of physical and digital infrastructure (Nelson and Ellis 2018), enabling their use in large-scale research in ecology, evolution, conservation biology, genomics, disease ecology, and various other fields (Ponder et al. 2001; Cook and Light 2019; Colella et al. 2021; Nachman et al. 2023). In parallel, field collection practices have evolved

substantially, with a single voucher specimen now comprising numerous specimen parts and digital assets, such as photographs and audio recordings. Additionally, the advent of genomics in biodiversity research has increased the value of collecting a variety of tissue types and preserving tissue duplicates and their nucleic acids (e.g., DNA and/or RNA) using different preservation methods (Salis et al. 2025, 2026). Furthermore, holistic specimen collection includes sampling parasites associated with host specimens, which are key resources for co-evolutionary studies, wildlife management, conservation, and pandemic preparedness (Cook et al. 2017; Bell et al. 2018; Galbreath et al. 2019; Colella et al. 2021; Paul et al. 2025). Modernizing field data collection and management would improve consistency and data integrity for the ever-growing set of physical and digital resources that form the "extended specimen" (Webster 2017; Hardisty et al. 2022).

For centuries, field data recording has relied on paper field catalogs (e.g., Hall 1962), whereas much of digitization has occurred post-collection. Attempts to record digitally in the field often rely on spreadsheet or proprietary form-development software, such as Claris FileMaker (https://www.claris.com/filemaker/). Paper catalogs are simple, robust in harsh field conditions, and independent of electrical power, but they are poorly suited for linking the many parts of natural history specimens, detailed information essential for genomic tissue collection, and their derivative digital data. Furthermore, paper notes require secondary data entry that is prone to transcription errors. For example, such errors are common during latitude and longitude data entry, and while detection tools exist to flag coordinate errors (e.g., https://www.geo-locate.org/; Zizka et al. 2019), they do not fix missing information in digitized records necessary for georeferencing the coordinates, such as uncertainty values (Marcer et al. 2021). Such transcription errors lead to inaccurate georeference data that can have downstream implications for our understanding of species' geographic ranges, and their use in ecological, biogeographical and conservation analyses (Maldonado et al. 2015; Bloom et al. 2018). Prior research has shown that geographic inaccuracies affect local estimates of species diversity more than does inaccurate taxonomy (Maldonado et al. 2015). Transcription errors can likewise affect taxon names, collection dates, and other metadata relevant for the biological question of interest (Garner et al. 2024), which creates an additional data-cleaning burden for researchers. Recording specimen data digitally at the point of collection eliminates the need for transcription and its associated risks.

As field expeditions become more collaborative, often involving researchers of multiple nationalities and with different data collection and management practices, sharing and reconciling multimodal field data risks information loss and corruption. Data standards such as Darwin Core (Wieczorek et al. 2012) and the Findability, Accessibility, Interoperability, and Reusability (FAIR) Principles (Wilkinson et al. 2016) provide a foundation for more consistent, interoperable, and reusable biodiversity data, but data management systems do not apply them consistently (Hackett et al. 2019). A broadly used data entry platform could bridge field data collection and museum curation without requiring institutions to replace their existing data management infrastructure. At the same time, advances in software frameworks, including

Flutter and Rust, have made it feasible to develop high-performance, cross-platform applications from a single codebase. Developing a specialized digital field catalog for NHCs would address the need to integrate and share data across research teams and institutions.

Beyond integrating data across teams and institutions, field cataloging tools help researchers use those data to guide ongoing sampling. Without readily accessible summaries of collected specimens and associated materials, gaps or imbalances across species, localities, sexes, or life stages may remain undetected until opportunities to address them have passed (see Daru et al. 2018; Cooper et al. 2019). Such information is particularly valuable when preservation resources are limited. An up-to-date inventory can help teams prioritize samples for liquid nitrogen and other preservation methods to optimize DNA quality while balancing representation across taxa and populations with adequate biological replication, minimizing impacts on downstream systematics and phylogenomic analyses (Pollock et al. 2002; Zwickl and Hillis 2002; Galbreath et al. 2019). These needs call for tools that combine reliable cataloging with real-time data summaries, work without internet connectivity, and remain accessible across devices and platforms in field settings with differing resources and infrastructure.

We developed a cross-platform field cataloging application, NAtural History Project Utilities (NAHPU), to streamline data collection and maintain data integrity from the time a specimen is collected to the moment it is deposited in natural history collections and uploaded to the collection management software, such as Arctos (Cicero et al. 2024), Specify (Beach 2018), Symbiota (Gries et al. 2014), and FileMaker (https://www.claris.com/filemaker/). NAHPU is designed to maximize information collected in the field and provide real-time data statistics to aid field decision-making. It supports both mobile and desktop platforms and is distributed freely through the platform's official stores for easy installation, updates, and maintenance.

Key Features

Data Entry

NAHPU is an offline, privacy-first field cataloging application. It supports common data-entry fields for mammalogy, ornithology, herpetology (beta), ichthyology (beta), and terrestrial and freshwater invertebrate zoology (beta). We are currently developing support for other disciplines, including paleontology, marine invertebrate zoology, and botany. NAHPU organizes records into projects. Each project can cover an expedition or a period of specimen preparation in the lab. After creating a Project, users can add Sites (field localities), Events (dates, times, and methods of site-specific collecting events), Specimens (organisms and other material collected during an Event), and Narratives (longer-form daily field notes). Each has an associated “form” on the NAHPU GUI, with a combination of standardized, user-customizable data-entry fields with media entries (see Figure 1a and e for Specimen and Narrative forms, respectively). The media entry field supports adding photos, audio, and videos accompanied by built-in gallery and media players, subject to the device’s available storage space (Figure 1b). NAHPU allows unlimited specimens and associated parts (Figure 1c), supporting rich data entries such as collection and

preservation method, preparators and other field personnel, and collection time for tissue samples. The Sites form also supports adding an unlimited number of coordinates manually, importing in common formats (CSV, TSV, JSON, KML, Shapefile, and GPX), or using the device's Global Positioning System (GPS) sensor. A built-in coordinate viewer (Figure 1d) supports basemap layers from Natural Earth (https://www.naturalearthdata.com/) and OpenStreetMap (https://www.openstreetmap.org/; styled by OpenFreeMap: https://github.com/hyperknot/openfreemap), as well as custom user maps in GeoJSON and WGS84 Shapefile formats. The Narrative field provides a user outlet for writing field notes with Markdown styling, such as italicizing scientific names, creating "daily activity summary" tables, and bolding key messages (Figure 1e). To simplify tissue curation after fieldwork, NAHPU provides a curation tool with a visual guide showing the location in a standard tissue storage box within a customizable storage bin or "cell" grid. The Curation form supports curation workflows using QR code scanning, tissue ID, or field ID (Figure 1f).

FAIR, Reproducibility, and Responsible Collecting

We designed NAHPU for reproducible data collection and responsible collecting practices, including collaborative fieldwork involving multiple institutions and local guides. Because institutions differ in their cataloging conventions and data requirements, NAHPU allows users to customize data-entry templates and controlled vocabularies through a graphical interface. This flexibility allows participants to follow their institutions' established practices while contributing records to the same expedition catalog. The Event forms also allow users to associate specimens with all individuals involved in their collection, ensuring that contributions from local guides and other team members are documented.

NAHPU supports FAIR data practices through record identifiers, shareable vocabularies, and standardized exports. The application assigns each project, specimen, and associated materials record a persistent identifier using Universally Unique Identifier (UUID) version 4 (generated using cryptographic random bits rather than the user's hardware information) to ensure each has a unique, privacy-protective identification code. By default, specimen parts use the specimen identifier, but users can add a UUID for each part or scan an institutional QR code associated with tissue tube labels. To maintain the existing field-collection practice, NAHPU includes an auto-incrementing field number for each new specimen, supporting the typical collector initials + number format and a more flexible prefix-number-suffix format (called project identifier in NAHPU). We also work toward internationalization, and have thus made NAHPU a multilingual application in English, Spanish, Brazilian Portuguese, and Indonesian, with the intention of adding support for more languages.

Data Statistics, Export, and Customization

The NAHPU environment features a growing list of data statistics and rich data export and sharing capabilities (Figure 1g and h). The statistics provide a text summary of the data (e.g., counts of specimens, parts, and sites) and interactive plots (e.g., specimens by methods and part

quantity by treatment) (Figure 1h). Project data can be exported as a Darwin Core Archive (DwC-A), Darwin Core Data Package (DwC-DP), NAHPU Data Package (NAHPU DP), PDF, or user-customizable tabular formats. NAHPU DP expands Darwin Core fields to include unmatched NAHPU fields and users' custom fields. It uses a standardized, Frictionless Data format (Fowler et al. 2018) for interoperability with other applications, including data analysis in Python and R using official or community-developed Frictionless libraries. The PDF export features include a built-in What You See Is What You Get (WYSIWYG) template editor with rich text formatting support, allowing users to create customizable document layouts for various PDF export and printing purposes, such as tissue and specimen labels, record archival, and expedition reports. Controlled vocabularies, export presets, project identity, and individual site and event records can be exported and imported as JSON files or scanned as a QR Code if the data fit the size limit (up to 2,953 bytes or English text characters). Users need to enter them only once for all devices used in cataloging, future projects, and collaboration. Users can also export all data summary, coordinates, and media files separately for use in external applications. Image exports support resizing and conversion to WEBP, PNG, and JPEG formats for use in presentation slides, websites, or field reports.

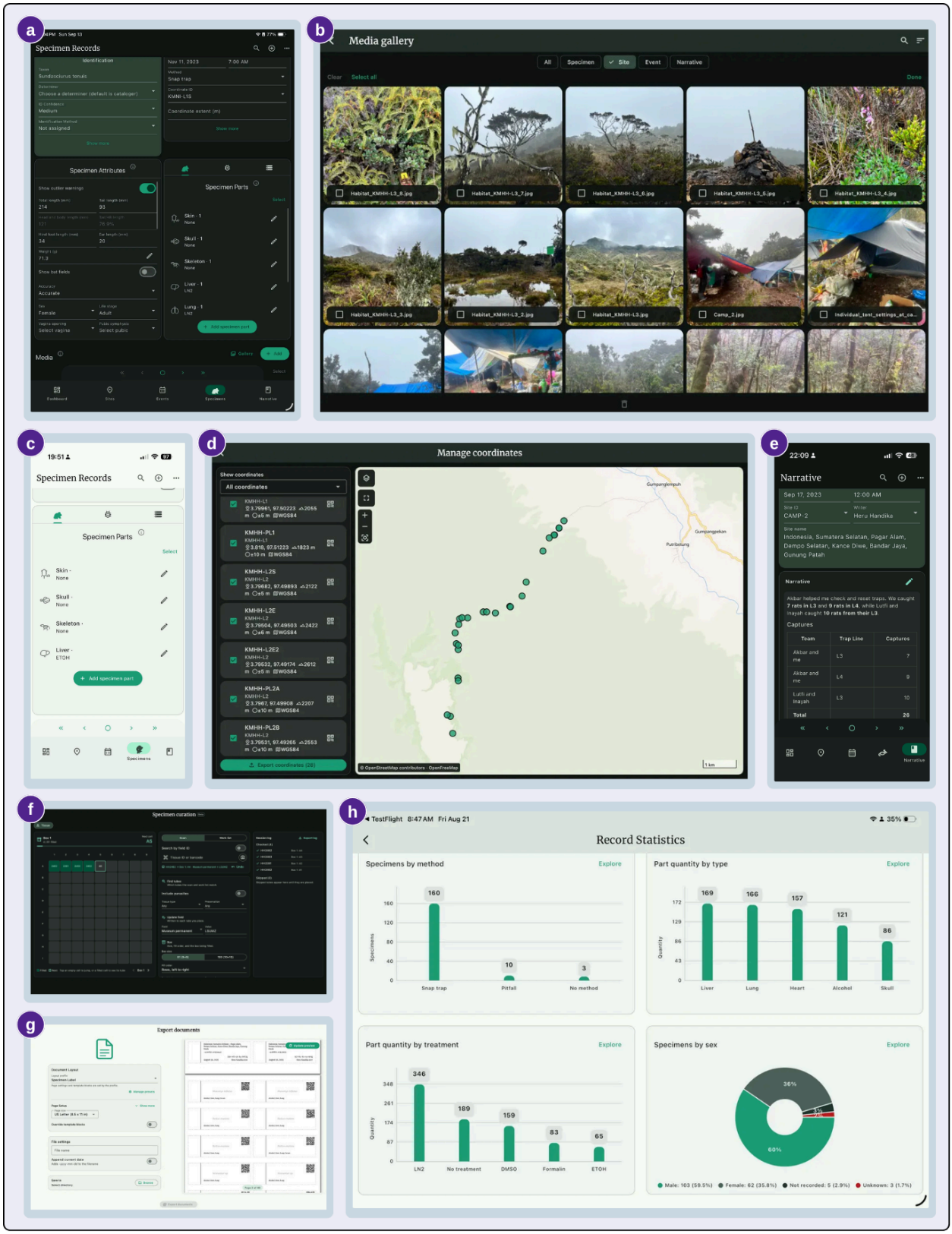


Figure 1. Screenshots of NAHPU v1.2.1 interface on various supported screen sizes (e.g., smart phone, tablet, laptop) using dark or light themes, and showcasing some of the features: (a) A

specimen form showing mammalogy format, (b) media gallery to browse through images attached to records, (c) an interface to add and manage specimen parts, (d) a coordinate management interface with a built-in map viewer, (e) a narrative field interface displaying a Markdown-rendered view of field journal, (f) a tissue curation form showing box location with guided visuals to enter the box and cell location, (g) a document export tool feature with a built-in PDF viewer, and (h) a graphical summary of the data updated in real time. NAHPU interfaces support both touch and mouse-click interaction. The a, g, and h panels are the interface layout on tablets; c and e are on phones, and b, d, f are on laptops and larger screens. All the screenshots show a subset of field data from a joint four-month expedition to three mountains in Sumatra, Indonesia, undertaken by the Indonesian National Research and Innovation Agency (BRIN) and the Louisiana State University Museum of Natural Science.

Workflows

NAHPU supports several field-specimen preparation workflows, which include parallel cataloging and preparation of specimens by different users from a shared site (Workflow A), parallel cataloging and preparation of specimens from different users working at separate sites (Workflow B), and assembly-line specimen preparation with centralized data entry by a single user on a single device (Workflow C) (Figure 2). Workflow A is commonly found in ornithology, terrestrial and freshwater invertebrate zoology, and volant mammalogy fieldwork. Specimens usually come from a shared mistnet, pitfall trap, Malaise trap, or harp trap. Each researcher works independently and records the specimens they sort and prepare on their own device. Workflow B is common in non-volant mammalogy fieldwork, where each team member independently runs trap lines and prepares and catalogs specimens. Workflow C is common in some herpetology, ichthyology, and mammalogy fieldwork, where each team member has a different role in specimen preparation (e.g., tissue collection or live specimen photography) and data entry is centralized in a single field catalog. For ichthyology and invertebrate zoology, in addition to single-individual entry, where each field identifier is an individual specimen, NAHPU also supports lot-based cataloging, where a single record contains multiple specimens of one species in a single lot. Users can also add unique tag identifiers (e.g., tissue tag number), individual measurements (e.g., standard length), and features, assigning specimen parts and parasite records to an individual fish or invertebrate, associated with their tag identifier.

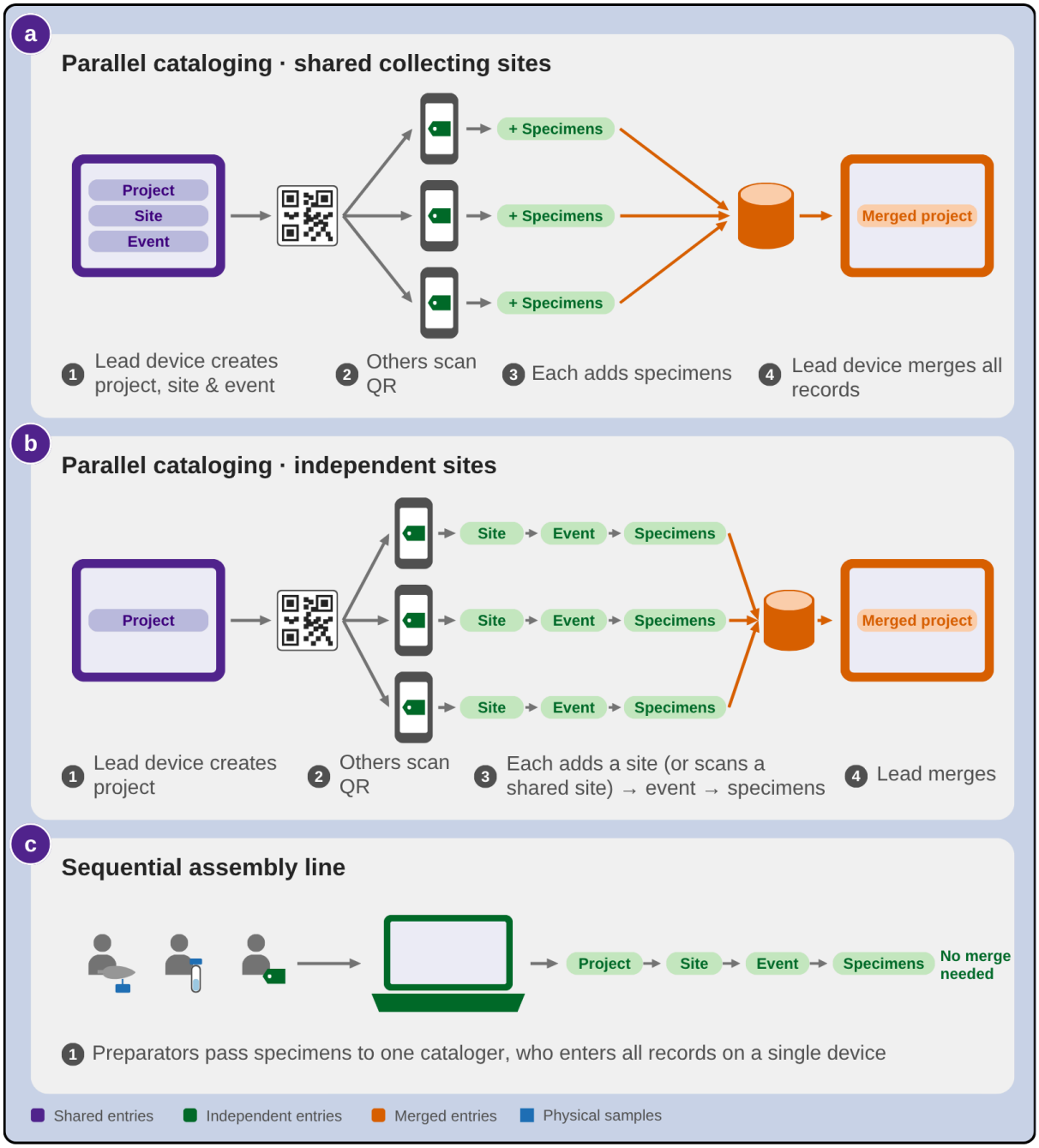


Figure 2. Typical field data recording workflows supported by NAHPU: (a) Specimens originate from a shared site and collecting event, such as bird specimens from a mistnetting check, while multiple team members prepare and cataloge the specimens independently using their own device; (b) team members run their own trapline and independently prepare and catalog the specimens, such as in small non-volant mammal collecting; (c) specimens are prepared through an assembly line where each individual team member performs a different role in specimen preparation, while a single team member is responsible for data entry, such as in herpetology specimen preparation. QR code sharing provides each device with identical project, site, and

event information. The device visuals are for illustration only; each workflow works on any NAHPU-supported platform.

Technical Implementation and Application Distribution

We designed NAHPU to support all major operating systems and Central Processing Unit (CPU) architectures, be extensible for future features, handle compute-intensive operations, and be easy to install. Its data schema includes exportable user configurations, shared entries, and project-specific entries (Figure 3a). Its interface adapts to the platform it runs on, supporting touch and mouse-click interaction. The application architecture is modular, divided into three main layers: the Dart+Flutter layers providing user interfaces (UIs), data access, and lightweight logic; the data layers in multiple data formats comprising project data, user configuration, and application configuration; and the Rust layers providing compute-intensive operations and reusable functions for various features, such as data conversion and compression, spatial tools, and PDF exports (Figure 3b). The UI leverages cross-platform Material Design 3 (M3; https://m3.material.io/) via the official Flutter M3 library, implemented with NAHPU-specific design tokens for consistent design throughout the application. The UI design follows the Web Content Accessibility Guidelines (WCAG) 2.2 Level AA for text (4.5:1 contrast ratio) and meets a 3:1 contrast ratio for graphics. The main project data is stored in a relational SQLite database (https://www.sqlite.org/), while user configurations use a key-value redb database (https://crates.io/crates/redb). Database schema changes are performed through migrations for backward compatibility. The Rust addition expands access to external, high-quality libraries beyond those available for Flutter and provides application programming interfaces (APIs) for use in other Rust applications. For example, the Rust API leverages the Typst engine (https://github.com/typst/typst), a powerful typesetting system comparable to LaTeX (https://latex-project.org/), to generate field note PDFs, and maps canvas pixel positions set in the Flutter-based Template Editor to the Typst system (Figure 3c and d). The APIs are also available as Rust packages distributed through the official Rust Package Registry (https://crates.io/).

The NAHPU application compiles into binaries for Android, iOS, iPadOS, macOS, Windows, and Linux, supporting x86_64 and ARM64 CPU architectures. We distribute it through the official stores for supported operating systems: the Apple App Store, Microsoft Store, Google Play Store, and Canonical Snap Store. For Linux, we also provide separate releases on GitHub in Debian format (.deb) for Debian and Ubuntu families, Red Hat Package Manager (.rpm) for Red Hat Enterprise Linux (RHEL) and Fedora families, and compressed TAR (.tar.gz) for other Linux distributions. All releases through official stores undergo machine and/or human expert reviews provided by the stores, follow store guidelines for application distribution and privacy-protection practices, and provide single-click installation and automatic updates. For the GitHub releases, we provide a script to download and verify application signature integrity and install with a single command, with or without root access.

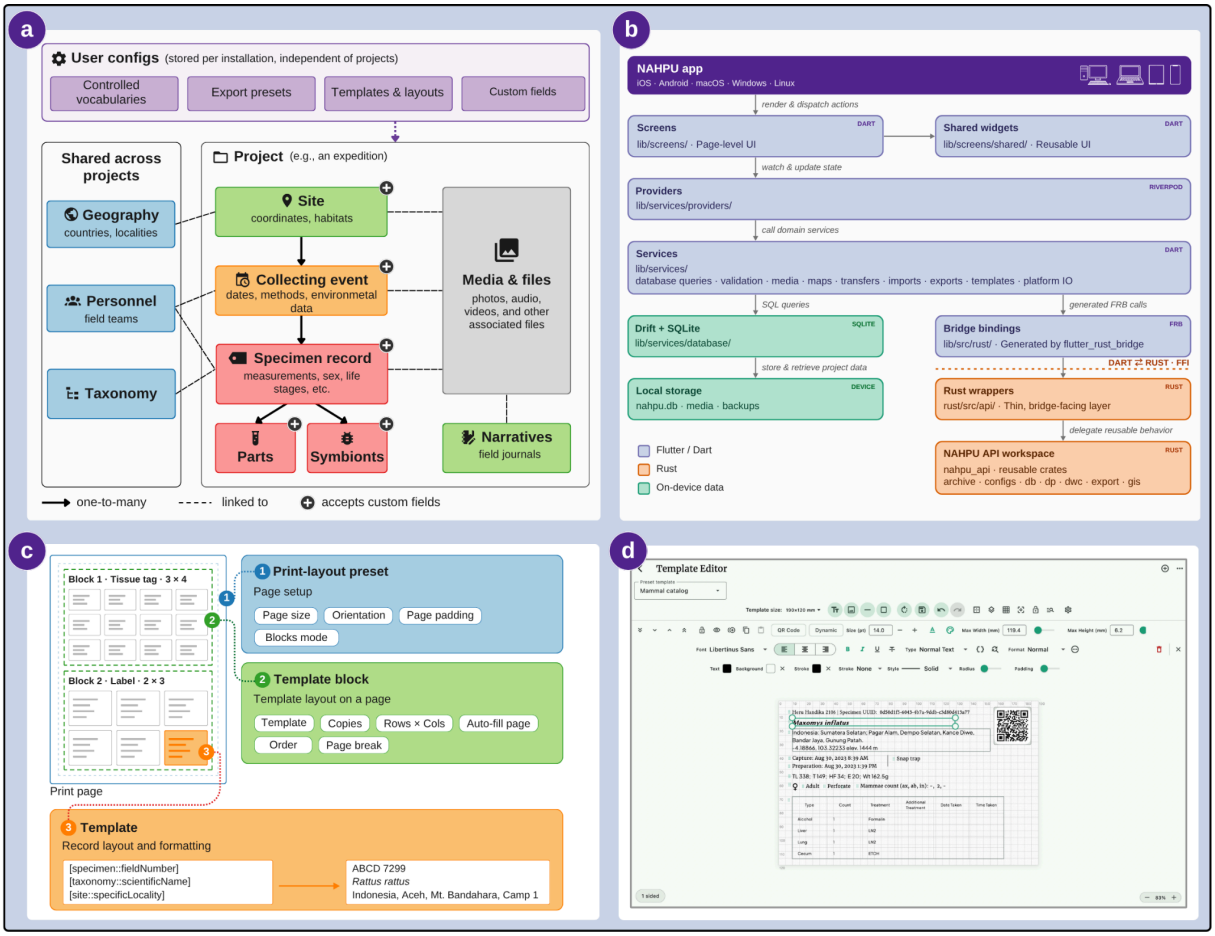


Figure 3. NAHPU modular architecture and document export support showing: (a) data association among user configurations, shared entries, and project-specific entries, (b) a high-level schema of the application architecture showing each layer of the application code, (c) a diagram of print layout showing how the preset, template block, and template fit together in a single-page layout, and (d) a screenshot of the WYSIWYG template editor on a large screen providing flexibility for users to design their own template. The document export system supports exporting PDFs of all project records for various purposes, such as record archival, specimen tags, and tissue labels.

Project Governance, Interdisciplinary Collaboration, and Training

We (the authors) currently govern the NAHPU project to minimize maintenance costs, provide interdisciplinary training, and offer a platform for implementing emerging software technologies to support biodiversity research. We began NAHPU development at Louisiana State University Museum of Natural Science (LSUMNS) and pushed our first commit to GitHub on 18 July 2022 (https://github.com/nahpu/nahpu/; commit #98cdcd9). Since then, we have expanded the project's collaboration and open-source coding practices. NAHPU currently represents a collaboration among curators, collection managers, and field biologists from seven natural

history museums and six universities. In addition, since August 2025, we have collaborated with the Human Augmented Analytics Group (HAAG) at Georgia Tech through the MENTOR-CS program (Shi et al. 2025), providing a platform for students in the Master of Science in Computer Science (OMSCS) to contribute to NAHPU as an interdisciplinary biodiversity science project. In June 2026, we expanded our NAHPU collaboration to support contributions from University of Florida undergraduate students in biology and computer science. Code contributors can work on GitHub issues or propose and develop independent projects to expand NAHPU features through collaboration with field biology team members.

Field Expeditions Using NAHPU

During its four years of development, several research teams tested NAHPU in seven field expeditions in Southeast Asia, South America, and Mexico, and six data-entry sessions from previous fieldwork in Southeast Asia and South China (Table 1), using a variety of devices, including phones, tablets, and laptops. The teams relied on power banks for expeditions without access to continuous electricity or on-site generators. A single iPad Mini 6 (if used exclusively for cataloging) along with a 25,000 mAh power bank could usually last three weeks. Most tablets were used with an external keyboard to speed up data entry. Data backup practices varied across teams. In expeditions in Brazil, we had access to satellite internet and backed up and integrated data daily. On the Sumatra field expedition, we cross-shared NAHPU data among cataloging devices using offline data-sharing features (e.g., AirDrop for Apple devices) and regularly stored it on a shockproof, waterproof flash drive for additional backup, and installed NAHPU on a personal phone as a secondary backup device. The flash drive also facilitates offline data sharing between Android and Apple devices. We uploaded the data to cloud services when we returned to a nearby town with internet and electricity access. We also tested NAHPU's features using a dataset from previous fieldwork on bats in karst habitat (i.e., caves and limestone forest), supplemented by specialized bat data including echolocation data forms, morphological photographs, tissue (wing biopsy punch), hair, and swab samples. Compared with former field cataloging approaches, NAHPU digital field catalogs enabled faster turnaround for exporting specimen records, summarizing captures for field reports, and accessioning them to the collection.

Table 1. Field expeditions using NAHPU for specimen cataloging, showing fieldwork dates, counts of specimens and specimen parts, and devices used for cataloging in the field and post-fieldwork. All team members cataloged all specimens exclusively using NAHPU (exclusive); not all collectors used NAHPU (partial); specimen data entered in paper catalogs first, then NAHPU (paper-first); or data entered into NAHPU after fieldwork (post-fieldwork). Asterisks indicate expeditions without on-site electricity.

Location	Discipline	Years	Duration (days)	Specimens/parts	Use	Devices

Mt. Gede, West Java, Indonesia*	Mammalogy	2023	33	196/846	partial	Lenovo Tab P11 Pro
Mts. Patah, Bandahara, Kemiri, Sumatra, Indonesia*	Mammalogy	2023	115	993/3605	exclusive	2 x iPad Mini 6, iPad Mini 4
Borneo, Indonesia/Malaysia*	Mammalogy	2024-2026	79	>500/>2500	exclusive	6 x iPad Mini 6
Mt. Kalatungan, Bukidnon, Phillippines*	Mammalogy	2025	20	468/1948	exclusive	iPad 6 and Samsung Galaxy A9
Cordillera Azul National Park, Peru	Ornithology	2023	38	1061	paper-first	MacBook Pro
Coari, Amazonas, Brazil	Ornithology	2025	30	604	paper-first	iPads and Android tablets
Mexico	Ichthyology	2024	13	1,670/351	paper-first & post-fieldwork	MacBook Pro
Yunnan Karst, Yunnan Province. China	Mammalogy	2017	60	258	post-fieldwork	iPad Air, iPhone SE, iPhone 13
Myanmar Karst	Mammalogy	2017	10	41	post-fieldwork	iPad Air, iPhone SE, iPhone 13
Yunnan Karst, Yunnan Province, China	Mammalogy	2018	60	144	post-fieldwork	iPad Air, iPhone SE, iPhone 13
Cat Ba Karst, Vietnam	Mammalogy	2018	10	88	post-fieldwork	iPad Air, iPhone SE, iPhone 13
Yunnan Karst, Yunnan Province, China	Mammalogy	2018	60	199	post-fieldwork	iPad Air, iPhone SE, iPhone 13
Sumatran Karst, Sumatra Island, Indonesia.	Mammalogy	2024	60	605/1117	post-fieldwork	iPad Air, iPhone SE, iPhone 13

Discussion

Specimen collection remains essential because non-lethal sampling methods (e.g., photographs and audio recordings) cannot substitute for the morphological, genetic, parasitological, and other data provided by whole specimens (Rocha et al. 2014; Thomson et al. 2018). We developed NAHPU to optimize natural history data collection for the extended specimen era by digitizing and linking all specimen data from the moment of their creation. In recent decades, natural history collections (NHCs) have expanded beyond physical specimens to include genetic, morphological, anatomical, parasite, ecological, geographic, multimedia, and field journal data (Lendemer et al. 2020; Hardisty et al. 2022; Sigwart et al. 2025; Del-Rio et al.). The value of these collections lies in the network of linked data around a voucher specimen (Lendemer et al. 2020; Hardisty et al. 2022). However, the network is usually assembled retrospectively, in some cases years later after the specimens and associated data have been collected, through digitization of the catalogs and re-linking of derivatives. NAHPU instead creates the linkages at the moment a specimen is collected, when they are cheapest and most reliable. Specimen parts, tissues, parasites, and derivative digital assets (photos/audio/videos) are attached directly to the voucher records. We include data summaries to support field decision-making, manage preservation resources, and check for balanced sampling of individuals or sex per population (Cooper et al. 2019). Ongoing work on outlier detection for specimen measurements flags potential errors before specimens are permanently prepared and characters become inaccessible for re-measurement. In environments where using an electronic device is impractical, such as in aquatic habitats, NAHPU could replace the spreadsheet entry step of this workflow, adding data-linkage and summary features while keeping the typical workflow for this environment.

We developed NAHPU features through experience working on complex and logistically challenging modern field expeditions (Table 1). We initially designed NAHPU based on a digital mammalogy field catalog developed by Kevin and Karen Rowe (Museums Victoria) in Claris FileMaker with support limited to Apple devices. The Rowe and Esselstyn labs (LSUMNS) used it for expeditions in Sulawesi, Java, and Sumatra, Indonesia, and Borneo, Malaysia, from 2011 until 2019. This catalog introduced a collecting-event section (similar to Hackett et al. 2019) to avoid duplicating collection-method and event entries for every specimen and to track the personnel involved, linking each specimen directly to its collectors, including local guides and local undergraduate students often involved in capturing it. NAHPU adapted and expanded its capability by adding support for tracking environmental data, customizable for both terrestrial and freshwater taxa, with support for custom fields, file attachment, and event-specific media (as of NAHPU v1.2.0; Figure 3a). This improves data linkage compared to a typical paper catalog format, where each catalog page is limited to sites and specimen records, while environmental data, personnel records, and digital assets (if recorded) often appear in separate notes (Hall 1962; Remsen 1977). Backing up paper catalogs in the field requires photographing each page, manually copying pages, or hand-entering data into a spreadsheet, whereas NAHPU backup only

takes a few clicks. It supports light export (< 1 megabyte for 1,000 records), simplifying uploads of essential records (excluding media) to cloud storage when internet access is limited.

NAHPU v1.2.0 added a tissue curation tool with QR code scanning and a visual interface that matches standard tissue box layouts to help organize tissue samples, informed by field expeditions in the Amazon and Sumatra that yielded over a thousand tissues in a single expedition. This feature also paves the way for organizing fossils collected in the field as we work towards supporting paleontology fieldwork in future releases, facilitating the enhancement of FAIR data practices (Long-Fox 2026), specimen digitization efforts (Boyer et al. 2016; Karim et al. 2016; Allmon et al. 2018; Blackburn et al. 2024; Bishop et al. 2025) and metadata standardization (Bishop et al. 2025) across paleontology NHCs and research groups. Together, NAHPU features support responsible, reproducible field expeditions while improving specimen organization.

NAHPU complements established collection management software (CMS), such as Specify (Beach 2018, 2019), Arctos (Cicero et al. 2024), Symbiota (Gries et al. 2014), and EMu (https://www.axiell.com/solutions/product/emu/). It fills a gap by streamlining data transfer from fieldwork to the museum database, supporting Darwin Core (DwC) exports for bulk import into CMS. Users can also create tabular export presets that match their CMS import columns. A customizable PDF export with a GUI template editor provides a simple solution for archiving field records in print. Inspired by Arctos-controlled vocabularies to ensure a consistent ontology, NAHPU supports user-customizable vocabularies with default values from Arctos and terminology used in published work, such as Lutz et al. (2017) and Galbreath et al. (2019) for parasites. While DwC facilitates information exchange, the standard adds terms only where shared demand has been demonstrated (Wieczorek et al., 2012). NAHPU allows users to define and share their own custom fields through the NAHPU data package in a flexible, Frictionless Data format (Fowler et al. 2018). While standardization matters for interoperability and sharing across collection platforms, it also has practical and theoretical limits in capturing the complexity of characters across the tree of life (Sterner et al. 2023). NAHPU balances this by providing a standard format and customization that users can create and share for reuse and reproducibility in a few clicks.

NAHPU installation and maintenance are like everyday software, with single-click installation and automatic updates provided by official app stores. Official app stores also add layers of testing and human review for security, safety, privacy, and performance, and they require maintaining compliance with evolving operating systems, which independent releases often lack. Because of its cross-platform approach from a single codebase enabled by Flutter and Rust, required fixes by stringent store reviews, such as Apple's App Store (review in McIlroy et al. 2016), carry over to other platforms. Its simple installation and compliance with modern software distribution make NAHPU broadly applicable for collection use at any scale. With continued support and upkeep, collections may discover NAHPU serves as collection management beyond its original intent for field data management, such as offline backup for large collections with established CMS or as the primary CMS for small collections that need a

low-cost digitization solution. For the latter, third-party cloud storage, such as Dropbox, Google Drive, and OneDrive, can back up NAHPU databases at little to no cost, depending on collection size. The expanded network of GBIF-registered organizations, supported by GBIF's Publishing Toolkit (IPT; Robertson et al., 2014) or other collection portals, such as Symbiota (https://symbiota.org/; Gries et al. 2014), could enable global data sharing without the laboratory maintaining its own database server, while data repository platforms, such as Zenodo (https://zenodo.org/), could also facilitate data sharing with permanent Digital Object Identifiers (DOIs).

NAHPU uses SQLite to store project data. This database engine is reliable, secure, widely used, supports up to 20 trillion rows of data entries, and has been around for more than two decades (https://www.sqlite.org/; Gaffney et al. 2022). SQLite projects are easy to compile and archive on cloud storage, adding a secondary layer of data security. If a collection decides to move away from NAHPU in the future, users will retain access to their records through widely available SQLite-compatible tools. The NAHPU data package also provides another complete data access option for spreadsheet workflows. In addition, to mitigate loss of application access and support reproducibility, we have kept the entire NAHPU codebase version-controlled and open-source since the first commit, including supporting software, evolving database schemas, and detailed documentation of its underlying architecture and design choices. The public code, database schemas, and documentation make it easier for new contributors to maintain NAHPU. Because NAHPU runs on users' hardware, increased adoption does not increase compute costs for the maintainers, and users can set up custom fields without coding or requiring maintainers to code them manually, keeping costs of using the application and maintenance low. The cost to release to official stores is shared with SEGUL (Handika and Esselstyn 2024) and the Mammal Diversity Database app (Burgin et al. 2025), at USD 99 per year at the time of writing this manuscript, at no cost to users. NHCs often face persistent financial constraints (Kemp 2015). The NAHPU software development approach mitigates the risk for users and maintainers.

Modern NHCs depend on effective data management, computational methods, and interdisciplinary collaboration, and accessible research software is an essential educational resource (Lendemer et al. 2020; Ellwood et al. 2026). Beyond field collection, NAHPU supports training in natural history collection and software development. It supports large numbers of devices; NAHPU v1.2.1 on Android alone supported 19,408 devices (Supplementary Table 1). Its features support the whole natural history pipeline from fieldwork to collection management, integrate into standard data collection and management workflows (Figure 3), and provide predictable data structures and GUI workflows that simplify training in specimen collection and management. Previous field or post-field testing of NAHPU capabilities in Brazil, China, Indonesia, Malaysia, Mexico, Myanmar, the Philippines, and Vietnam demonstrates NAHPU's flexibility for applied field data collection in a variety of geographic and cultural contexts. Since the release of NAHPU v1.2.0, we have also expanded language support beyond English by adding beta support for machine translation in Spanish, Brazilian Portuguese, and Indonesian, with a system in place for human reviews and updates. In addition, NAHPU has served as a

springboard for cross-institutional collaborations (with HAAG and other groups) and undergraduate and graduate software development training and mentorship (see Shi et al. 2025).

Limitations

Our deliberate design decision to develop NAHPU as an offline-focused application protects user information and minimizes cybersecurity risks for users. However, it also precludes automatic data synchronization support. Instead, we provide manual data merging with syncing-related conflict-resolution features. We recommend using NAHPU in the field to establish a primary authoritative device, so that merging conflicts, such as site information, can be resolved to a single primary reference source. To share project data with a global audience, the offline approach will also require users to upload NAHPU data to an online collection management system, such as Specify (Beach 2018), Arctos (Cicero et al. 2024), and Symbiota (Gries et al. 2014), GBIF IPT (Robertson et al. 2014) for GBIF-registered organizations, or online dataset repositories such as Zenodo (https://zenodo.org/). While we develop NAHPU from a single codebase for all supported platforms, different devices and operating systems provide different capabilities because the cross-platform framework abstracts over platforms that remain heterogeneous in their ecosystem and hardware capabilities (Biørn-Hansen et al. 2020). For example, camera access for QR code scanning is available only on mobile devices and macOS. We provided JSON import as an alternative.

Future Directions

NAHPU's low computing costs and direct, low-latency hardware access, compared with web-based, server-run software, allow flexibility to expand beyond collection management and basic statistics with minimal impact to recurring development costs. Future development will extend NAHPU from a data management application into an integrated data analytic platform, expanding its use for downstream research involving NHCs. New releases will soon integrate external resources such as the Catalogue of Life (Cachuela-Palacio 2006), GBIF, GenBank (https://www.ncbi.nlm.nih.gov/genbank/), and Crossref (https://www.crossref.org/), similar to BioCosmos species pages (https://lepiverse.com; Handika et al. 2026). This information will be available in the specimen forms, allowing users to identify and update outdated taxonomy, recognize gaps in specimen or tissue sampling, and discover relevant genomic and published resources. Advances in digital specimen analysis and automated curation offer opportunities for computational approaches to enhance the research value of NHCs (Hedrick et al., 2020). We plan to expand data linking for externally managed specimen-derivative datasets, such as photographs and genetic sequences, and implement more advanced statistical analysis features for morphological and genetic data. On-device integration of principal component and allometric analyses (Chan and Grismer 2025, similar to 2026) will expand outlier detection and facilitate the production of specimen morphospaces that update in real time during field collection. NAHPU, which shares an underlying architecture with SEGUL, offers a straightforward path to integrate SEGUL phylogenomics features for summary statistics, filtering, and genomic

sequence concatenation (Handika and Esselstyn, 2024), allowing users to connect genomic data to specimen metadata within NAHPU and prepare them for downstream analyses, such as phylogenomics.

Advances in computer vision and metric learning methods show potential to reveal patterns of adaptive convergence (e.g., in butterfly coloration) that conventional phenomic methods (such as manual phenotype scoring or geometric morphometrics) sometimes struggle to characterize (Hoyal Cuthill et al. 2019; Puissant et al. 2023, 2026; Lürig et al. 2024, 2025). Ongoing work on MILU (https://github.com/nahpu/milu) combines cross-platform frameworks and GPU-agnostic computing by using Burn (https://burn.dev/) and WGPU (https://github.com/gfx-rs/wgpu) to simplify specimen image analysis via on-device deep learning and metric learning. The feature will be optional within NAHPU to avoid disrupting the existing workflow for field cataloging, with potential use beyond computer vision analyses. Increased specimen digitization efforts have a demonstrated potential to increase access to and use of NHCs for a wide range of applications by scientists and the public alike (Hedrick et al. 2020). As new software technologies and analytical needs emerge, NAHPU's modular architecture and native access to hardware approaches offer opportunities to expand application capabilities to support specimen digitization and data analyses in ways that are accessible for evolutionary biologists regardless of technical and funding support.

Acknowledgements

We thank Kevin C. Rowe and Karen M. C. Rowe for developing a Claris FileMaker Pro digital field catalog that inspired NAHPU's initial interface design and user experience workflow. We express our gratitude to the following individuals for their feedback on the app features and usability: Alfath Fanidya, Endah D. Jayanti, Fajar Alfitrian, Lutfi S. Maulidiyah, Akbar F. Hilmi, Spenser J. Babb-Biernacki, Katherine Wolcott, Lucas T. Bezerra, Darwin Morales-Martínez, Diego Cueva, Edward L. Stanley, Jocelyn P. Colella, Jessie L. Williamson, Nicholas A. Mason, Katrina Derieg, Ellie Dripps, Alivia Hartz, Eleanor Hoeger, Gabriella Linsalata, Lazaro Lopez, John Mewherter, Fritz P. Marcano, Heidi Stevens, Marisa Surovy, Rose Wilhoyt, Litsa Wooten, Melissa T. R. Hawkins, and Lucas H. DeCicco. Charles R. Clark, Srijan Srivatsa, Ashwin Narayanan, and other members of the admin team at The Human-Augmented Analytics Group (HAAG), the Georgia Institute of Technology, provided logistical support and team management for interdisciplinary collaboration. The National Science Foundation (NSF DEB-1754393 and DEB-2244754) and the Alfred L. Gardner and Mark S. Hafner Mammalogy Fund provided financial support for the development. Additional funding from various sources supported contributors during the project (Arthur Porto and Caleb M. Gordon: NSF under Grant No. EAR-2422839, Jonathan Nations: NSF Postdoctoral Fellowship in Biology under Grant No. DBI-2010756). Louisiana State University, the University of Florida, and Georgia Institute of Technology provided the computing resources for software development.

Author Contribution

List of contributions:
Conceptualization, app data curation, funding acquisition, methodology, code writing, code reviewing, feature testing and validation, software documentation, documentation translation, visualization, manuscript writing

Heru Handika (conceptualization, app data curation, methodology, code writing, feature testing and validation, software documentation, code reviewing, visualization, manuscript writing), Andre E. Moncrieff (conceptualization, app data curation, methodology, feature testing and validation, software documentation, manuscript writing), Cody Henderson (methodology, code writing, code reviewing), Alain Dadaian (methodology, code writing), Austin S. Chipps (app data curation, feature testing and validation, manuscript writing), Nurul Inayah (app data curation, feature testing and validation), Ada Chornelia (conceptualization, feature testing and validation, documentation translation, manuscript writing), Natalia Pérez-Amaya (feature testing and validation, documentation translation), Roberta Canton (app data curation), Marco A. Rego (app data curation), Aakanksha T. Jaiswal (code writing, manuscript writing), Daeun D. Lee (code writing), Jia Yu (code writing), Ashmitha Julius (code, writing), Lawrence B. Gan (code writing), Dante R. Evangelista (code writing), Samantha L. Rutledge (methodology, feature testing and validation), Lauren Vonnahme (methodology, feature testing and validation), Caleb M. Gordon (methodology, feature testing and validation, software documentation, manuscript writing), Breanna Shi (conceptualization, funding acquisition, interdisciplinary collaboration), Diego J. Elías (methodology, feature testing and validation, manuscript writing), Emily M. Troyer (methodology, feature testing and validation, manuscript writing), Gustavo Silva de Miranda (methodology, feature testing and validation, manuscript writing), Jonathan Nations (conceptualization, methodology, feature testing, and validation, manuscript writing), Glaucia Del-Rio (conceptualization, methodology, feature testing and validation, manuscript writing), Gregory Thom (conceptualization, methodology, feature testing and validation, manuscript writing), Arthur Porto (conceptualization, methodology, interdisciplinary collaboration, funding acquisition, manuscript writing), Jacob A. Esselstyn (conceptualization, methodology, feature testing and validation, funding acquisition, manuscript writing). All authors reviewed and approved the final version of the manuscript.

The Use of Generative Large-Language Models

We maintain a human-in-the-loop approach to using generative artificial intelligence (AI) for software development and writing. We used several frontier models, including Anthropic's Claude Opus (v4.5 and later), OpenAI's GPT (v5 and later), and Google Gemini Flash, Pro, and Argon (v3 and later), to generate code (in commits since September 2025), graphics, and language edits for the documentation and manuscript. Humans reviewed AI usage and tested the application. AI-generated documentation translations are labeled as AI writing before human

revisions. Humans review all code pull requests and may use frontier models for secondary checks after merging the code. Living guidelines on using LLMs for NAHPU development are available here: https://nahpu.app/en/contributing/code/llm-assisted-development/ (https://github.com/nahpu/nahpu-docs).

App and Code Availability

This manuscript describes NAHPU up to v1.2.1. Code availability is as follows: the NAHPU app at https://github.com/nahpu/nahpu, the API at https://github.com/nahpu/nahpu_api, and the documentation at https://github.com/nahpu/nahpu-docs. The entire codebase is maintained in the organization's GitHub account (https://github.com/nahpu). NAHPU app can be downloaded through the official stores of supported devices: the Apple App Store (https://apps.apple.com/us/app/nahpu/id1661448584), the Google Play (https://play.google.com/store/apps/details?id=com.hhandika.nahpu), the Microsoft Store (https://apps.microsoft.com/detail/9nx3z8p3nsml), and the Canonical Snapcraft (https://snapcraft.io/nahpu)

References

Allmon W.D., Dietl G.P., Hendricks J.R., Ross R.M. 2018. Bridging the two fossil records: Paleontology's "big data" future resides in museum collections. Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making. Geological Society of America. p. 35–44.

Beach J. 2018. Specify Collections Consortium–Building Durable Infrastructure. Biodivers. Inf. Sci. Stand. 2:e26860.

Beach J. 2019. Biodiversity community cyberinfrastructure in the Specify Collections Consortium. Biodivers. Inf. Sci. Stand. 3.

Bell K.C., Carlson C.J., Phillips A.J. 2018. Parasite Collections: Overlooked Resources for Integrative Research and Conservation. Trends Parasitol. 34:637–639.

Biørn-Hansen A., Rieger C., Grønli T.-M., Majchrzak T.A., Ghinea G. 2020. An empirical investigation of performance overhead in cross-platform mobile development frameworks. Empir. Softw. Eng. 25:2997–3040.

Bishop B.W., Fox J., Gavel S.W.T., Chapin E.G., Kansa S. 2025. From ice cores to dinosaurs: physical collections managers' research data curation perceptions and behaviors. PLoS One. 20:e0317070.

Blackburn D.C., Boyer D.M., Gray J.A., Winchester J., Bates J.M., Baumgart S.L., Braker E.,

Coldren D., Conway K.W., Rabosky A.D., de la Sancha N., Dillman C.B., Dunnum J.L., Early C.M., Frable B.W., Gage M.W., Hanken J., Maisano J.A., Marks B.D., Maslenikov K.P., McCormack J.E., Nagesan R.S., Pandelis G.G., Prestridge H.L., Rabosky D.L., Randall Z.S., Robbins M.B., Scheinberg L.A., Spencer C.L., Summers A.P., Tapanila L., Thompson C.W., Tornabene L., Watkins-Colwell G.J., Welton L.J., oVert Project Team, Stanley E.L. 2024. Increasing the impact of vertebrate scientific collections through 3D imaging: The openVertebrate (oVert) Thematic Collections Network. Bioscience. 74:169–186.

Bloom T.D.S., Flower A., DeChaine E.G. 2018. Why georeferencing matters: Introducing a practical protocol to prepare species occurrence records for spatial analysis. Ecol. Evol. 8:765–777.

Boldgiv B., Lkhagva A., Edwards S., Stenseth N.C., Bayarsaikhan J., Altangerel D., Usukhjargal D., Dovchin B., Gombobaatar S., Batsaikhan N., Warinner C., Hart I., Galbreath K., Greiman S.E., Malaney J., Murdoch J.D., McLean B., DeWitte S.N., Manzitto-Tripp E., Chin K., Karim T.S., Simpson C., Stevens N.J., Dunnum J.L., Cook J.A., Taylor W.T.T. 2025. Global natural history infrastructure requires international solidarity, support, and investment in local capacity. Proc. Natl. Acad. Sci. U. S. A. 122.

Boyer D.M., Gunnell G.F., Kaufman S., McGeary T.M. 2016. Morphosource: Archiving and sharing 3-d digital specimen data. Paleontol. Soc. Pap. 22:157–181.

Burgin C.J., Zijlstra J.S., Becker M.A., Handika H., Alston J.M., Widness J., Liphardt S., Huckaby D.G., Upham N.S. 2025. How many mammal species are there now? Updates and trends in taxonomic, nomenclatural, and geographic knowledge. J. Mammal.

Cachuela-Palacio M. 2006. Towards an index of all known species: the Catalogue of Life, its rationale, design and use. Integr. Zool. 1:18–21.

Chan K.O., Grismer L.L. 2025. GroupStruct2: A user-friendly graphical user interface for statistical and visual support in species diagnosis. Syst. Biol.:syaf090.

Chan K.O., Grismer L.L. 2026. Extending GroupStruct2: a Bayesian and machine-learning framework for testing taxonomic hypotheses using morphometric data. Zookeys. 1276:125–138.

Cicero C., Koo M.S., Braker E., Abbott J., Bloom D., Campbell M., Cook J.A., Demboski J.R., Doll A.C., Frederick L.M., Linn A.J., Mayfield-Meyer T.J., McDonald D.L., Nachman M.W., Olson L.E., Roberts D., Sikes D.S., Witt C.C., Wommack E.A. 2024. Arctos: Community-driven innovations for managing natural and cultural history collections. PLoS One. 19:e0296478.

Colella J.P., Bates J., Burneo S.F., Camacho M.A., Carrion Bonilla C., Constable I., D'Elía G., Dunnum J.L., Greiman S., Hoberg E.P., Lessa E., Liphardt S.W., Londoño-Gaviria M., Losos E., Lutz H.L., Ordóñez Garza N., Peterson A.T., Martin M.L., Ribas C.C., Struminger B., Torres-Pérez F., Thompson C.W., Weksler M., Cook J.A. 2021. Leveraging natural history biorepositories as a global, decentralized, pathogen surveillance network. PLoS

Pathog. 17:e1009583.

Cook J.A., Edwards S.V., Lacey E.A., Guralnick R.P., Soltis P.S., Soltis D.E., Welch C.K., Bell K.C., Galbreath K.E., Himes C., Allen J.M., Heath T.A., Carnaval A.C., Cooper K.L., Liu M., Hanken J., Ickert-Bond S. 2014. Natural history collections as emerging resources for innovative education. Bioscience. 64:725–734.

Cook J.A., Galbreath K.E., Bell K.C., Campbell M.L., Carrière S., Colella J.P., Dawson N.G., Dunnum J.L., Eckerlin R.P., Fedorov V., Greiman S.E., Haas G.M.S., Haukisalmi V., Henttonen H., Hope A.G., Jackson D., Jung T.S., Koehler A.V., Kinsella J.M., Krejsa D., Kutz S.J., Liphardt S., MacDonald S.O., Malaney J.L., Makarikov A., Martin J., McLean B.S., Mulders R., Nyamsuren B., Talbot S.L., Tkach V.V., Tsvetkova A., Toman H.M., Waltari E.C., Whitman J.S., Hoberg E.P. 2017. The Beringian Coevolution Project: holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. Arct. Sci. 3:585–617.

Cook J.A., Light J.E. 2019. The emerging role of mammal collections in 21st century mammalogy. J. Mammal. 100:733–750.

Cooper N., Bond A.L., Davis J.L., Portela Miguez R., Tomsett L., Helgen K.M. 2019. Sex biases in bird and mammal natural history collections. Proc. Biol. Sci. 286:20192025.

Daru B.H., Park D.S., Primack R.B., Willis C.G., Barrington D.S., Whitfeld T.J.S., Seidler T.G., Sweeney P.W., Foster D.R., Ellison A.M., Davis C.C. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytol. 217:939–955.

Del-Rio G., Rego M.A., Baumgart S.L., Randall Z.S., Kratter A.W., Keriazes M., Doyle Z., Marcelino R.S., Handika H., Truong H.C., Schachner E.R. Integrating specimens' internal anatomy with external morphology in modern ornithological collections. Ibis (Lond. 1859).

Ellwood E.R., Nelson G., Zimkus B., Bates J., Cook J.A., Pandey J., Sessa E., Borths M., Goodwin J., Jennings D., Soltis P.S. 2026. From knowledge to action: Next steps for the natural science collections community. Bioscience.:biag018.

Fowler D., Barratt J., Walsh P. 2018. Frictionless data: Making research data quality visible. Int. J. Digit. Curation. 12:274–285.

Gaffney K.P., Prammer M., Brasfield L., Hipp D.R., Kennedy D., Patel J.M. 2022. SQLite: past, present, and future. Proceedings VLDB Endowment. 15:3535–3547.

Galbreath K.E., Hoberg E.P., Cook J.A., Armién B., Bell K.C., Campbell M.L., Dunnum J.L., Dursahinhan A.T., Eckerlin R.P., Gardner S.L., Greiman S.E., Henttonen H., Jiménez F.A., Koehler A.V.A., Nyamsuren B., Tkach V.V., Torres-Pérez F., Tsvetkova A., Hope A.G. 2019. Building an integrated infrastructure for exploring biodiversity: field collections and archives of mammals and parasites. J. Mammal. 100:382–393.

Garner B.H., Creedy T.J., Allan E.L., Crowther R., Devenish E., Kokkini P., Livermore L., Lohonya K., Lowndes N., Wing P., Vogler A.P. 2024. The taxonomic composition and

chronology of a museum collection of Coleoptera revealed through large-scale digitisation. Front. Ecol. Evol. 12:1305931.

Gries C., Gilbert E.E., Franz N.M. 2014. Symbiota - A virtual platform for creating voucher-based biodiversity information communities. Biodivers. Data J. 2:e1114.

Hackett R.A., Belitz M.W., Gilbert E.E., Monfils A.K. 2019. A data management workflow of biodiversity data from the field to data users. Appl. Plant Sci. 7:e11310.

Hall E.R. 1962. Collecting and preparing study specimens of vertebrates. University of Kansas.

Handika H., Esselstyn J.A. 2024. SEGUL: Ultrafast, memory-efficient and mobile-friendly software for manipulating and summarizing phylogenomic datasets. Mol. Ecol. Resour. 24:e13964.

Handika H., Lürig M.D., Nichtawitz K.L., Dobon O.M., Yarava V.S., Sobayo J.K., Shi B., Elliott M.J., Fortes J.A.B., Porto A. 2026. BioCosmos: computer vision and large language models for large-scale image-based biodiversity data and trait exploration. Available from https://ecoevorxiv.org/repository/view/15964/.

Hardisty A.R., Ellwood E.R., Nelson G., Zimkus B., Buschbom J., Addink W., Rabeler R.K., Bates J., Bentley A., Fortes J.A.B., Hansen S., Macklin J.A., Mast A.R., Miller J.T., Monfils A.K., Paul D.L., Wallis E., Webster M. 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the Internet. Bioscience. 72:978–987.

Hedrick B.P., Heberling J.M., Meineke E.K., Turner K.G., Grassa C.J., Park D.S., Kennedy J., Clarke J.A., Cook J.A., Blackburn D.C., Edwards S.V., Davis C.C. 2020. Digitization and the Future of Natural History Collections. Bioscience. 70:243–251.

Hoyal Cuthill J.F., Guttenberg N., Ledger S., Crowther R., Huertas B. 2019. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. Sci. Adv. 5:eaaw4967.

Karim T.S., Burkhalter R., Farrell Ú.C., Molineux A., Nelson G., Utrup J., Butts S.H. 2016. Digitization workflows for paleontology collections. Palaeontologia Electronica. 19:1–14.

Kemp C. 2015. Museums: The endangered dead. Nature. 518:292–294.

Lendemer J., Thiers B., Monfils A.K., Zaspel J., Ellwood E.R., Bentley A., LeVan K., Bates J., Jennings D., Contreras D., Lagomarsino L., Mabee P., Ford L.S., Guralnick R., Gropp R.E., Revelez M., Cobb N., Seltmann K., Aime M.C. 2020. The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. Bioscience. 70:23–30.

Long-Fox B.L. 2026. Open and FAIR data practices in palaeontology: guidance for authors at *Historical Biology*. Hist. Biol. 38:3715–3720.

Lürig M.D., Di Martino E., Porto A. 2024. BioEncoder: A metric learning toolkit for

comparative organismal biology. Ecol. Lett. 27:e14495.

Lürig M.D., Shirai L.T., Mota L., Willmott K.R., Freitas A.V.L., Porto A. 2025. Aposematic color patterns are the dominant axis of phenotypic diversification in Nymphalid butterflies. bioRxiv.:2025.09.30.678834.

Lutz H.L., Tkach V.V., Weckstein J.D. 2017. Methods for specimen-based studies of avian symbionts. In: Webster M.S., editor. The Extended Specimen: Emerging Frontiers in Collections-based Ornithological Research. Boca Raton, FL: CRC Press. p. 157–183.

Maldonado C., Molina C.I., Zizka A., Persson C., Taylor C.M., Albán J., Chilquillo E., Rønsted N., Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases?: Species diversity and distribution in the era of Big Data. Glob. Ecol. Biogeogr. 24:973–984.

Marcer A., Haston E., Groom Q., Ariño A.H., Chapman A.D., Bakken T., Braun P., Dillen M., Ernst M., Escobar A., Fichtmüller D., Livermore L., Nicolson N., Paragamian K., Paul D., Pettersson L.B., Phillips S., Plummer J., Rainer H., Rey I., Robertson T., Röpert D., Santos J., Uribe F., Waller J., Wieczorek J.R. 2021. Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. Divers. Distrib. 27:564–567.

McIlroy S., Ali N., Hassan A.E. 2016. Fresh apps: an empirical study of frequently-updated mobile apps in the Google play store. Empir. Softw. Eng. 21:1346–1370.

Nachman M.W., Beckman E.J., Bowie R.C.K., Cicero C., Conroy C.J., Dudley R., Hayes T.B., Koo M.S., Lacey E.A., Martin C.H., McGuire J.A., Patton J.L., Spencer C.L., Tarvin R.D., Wake M.H., Wang I.J., Achmadi A., Álvarez-Castañeda S.T., Andersen M.J., Arroyave J., Austin C.C., Barker F.K., Barrow L.N., Barrowclough G.F., Bates J., Bauer A.M., Bell K.C., Bell R.C., Bronson A.W., Brown R.M., Burbrink F.T., Burns K.J., Cadena C.D., Cannatella D.C., Castoe T.A., Chakrabarty P., Colella J.P., Cook J.A., Cracraft J.L., Davis D.R., Davis Rabosky A.R., D'Elía G., Dumbacher J.P., Dunnum J.L., Edwards S.V., Esselstyn J.A., Faivovich J., Fjeldså J., Flores-Villela O.A., Ford K., Fuchs J., Fujita M.K., Good J.M., Greenbaum E., Greene H.W., Hackett S., Hamidy A., Hanken J., Haryoko T., Hawkins M.T.R., Heaney L.R., Hillis D.M., Hollingsworth B.D., Hornsby A.D., Hosner P.A., Irham M., Jansa S., Jiménez R.A., Joseph L., Kirchman J.J., LaDuc T.J., Leaché A.D., Lessa E.P., López-Fernández H., Mason N.A., McCormack J.E., McMahan C.D., Moyle R.G., Ojeda R.A., Olson L.E., Kin Onn C., Parenti L.R., Parra-Olea G., Patterson B.D., Pauly G.B., Pavan S.E., Peterson A.T., Poe S., Rabosky D.L., Raxworthy C.J., Reddy S., Rico-Guevara A., Riyanto A., Rocha L.A., Ron S.R., Rovito S.M., Rowe K.C., Rowley J., Ruane S., Salazar-Valenzuela D., Shultz A.J., Sidlauskas B., Sikes D.S., Simmons N.B., Stiassny M.L.J., Streicher J.W., Stuart B.L., Summers A.P., Tavera J., Teta P., Thompson C.W., Timm R.M., Torres-Carvajal O., Voelker G., Voss R.S., Winker K., Witt C., Wommack E.A., Zink R.M. 2023. Specimen collection is essential for modern science. PLoS Biol. 21:e3002318.

Nelson G., Ellis S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374:20170391.

Page L.M., MacFadden B.J., Fortes J.A., Soltis P.S., Riccardi G. 2015. Digitization of

biodiversity collections reveals biggest data on biodiversity. Bioscience. 65:841–842.

Paul D.L., Thompson C.W., Arroyo L., Nuñez G.B., Castro-Arellano I., Colella J.P., Cook J.A., Cove M.V., Dearborn J., de la Sancha N.U., Derieg K., Dunnum J.L., Ferguson A.W., Fitzgerald K., Foat C., Frank L.E., Galbreath K., Gonzalez D.F., Hawkins R.K., Herrera T., Hey A., Hope A.G., Hornsby A.D., Ibanez D. IV, Keller O., Krejsa D., Mahoney K., Malaney J., McDonough M.M., McElrath T.C., Mathis V., Mayfield-Meyer T., Mosley I., Ordóñez-Garza N., Phelps K., Platt R.N. II, Pletcher J., Rinsland H.H., Rupprecht C.E., Speer K.A., Steger L., Upham N., White R.A. III, Wiens B.J., Wooten L., Soltis P.S. 2025. Harnessing natural history collections for collaborative pandemic preparedness. Bioscience.:biaf035.

Pollock D.D., Zwickl D.J., McGuire J.A., Hillis D.M. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Syst. Biol. 51:664–671.

Ponder W.F., Carter G.A., Flemons P., Chapman R.R. 2001. Evaluation of museum collection data for use in biodiversity assessment. Conserv. Biol. 15:648–657.

Puissant A., Chotard A., Condamine F.L., Debat V., Llaurens V. 2026. The coevolution of colour patterns and hindwing shapes on a large phylogenetic scale reveals predation-driven adaptive syndromes in swallowtail butterflies. Ecol. Lett. 29:e70303.

Puissant A., Chotard A., Condamine F.L., Llaurens V. 2023. Convergence in sympatric swallowtail butterflies reveals ecological interactions as a key driver of worldwide trait diversification. Proc. Natl. Acad. Sci. U. S. A. 120:e2303060120.

Remsen J.V. Jr. 1977. On taking field notes. American Birds. 31:946–953.

Robertson T., Döring M., Guralnick R., Bloom D., Wieczorek J., Braak K., Otegui J., Russell L., Desmet P. 2014. The GBIF integrated publishing toolkit: facilitating the efficient publishing of biodiversity data on the internet. PLoS One. 9:e102623.

Rocha L.A., Aleixo A., Allen G., Almeda F., Baldwin C.C., Barclay M.V.L., Bates J.M., Bauer A.M., Benzoni F., Berns C.M., Berumen M.L., Blackburn D.C., Blum S., Bolaños F., Bowie R.C.K., Britz R., Brown R.M., Cadena C.D., Carpenter K., Ceríaco L.M., Chakrabarty P., Chaves G., Choat J.H., Clements K.D., Collette B.B., Collins A., Coyne J., Cracraft J., Daniel T., de Carvalho M.R., de Queiroz K., Di Dario F., Drewes R., Dumbacher J.P., Engilis A. Jr, Erdmann M.V., Eschmeyer W., Feldman C.R., Fisher B.L., Fjeldså J., Fritsch P.W., Fuchs J., Getahun A., Gill A., Gomon M., Gosliner T., Graves G.R., Griswold C.E., Guralnick R., Hartel K., Helgen K.M., Ho H., Iskandar D.T., Iwamoto T., Jaafar Z., James H.F., Johnson D., Kavanaugh D., Knowlton N., Lacey E., Larson H.K., Last P., Leis J.M., Lessios H., Liebherr J., Lowman M., Mahler D.L., Mamonekene V., Matsuura K., Mayer G.C., Mays H. Jr, McCosker J., McDiarmid R.W., McGuire J., Miller M.J., Mooi R., Mooi R.D., Moritz C., Myers P., Nachman M.W., Nussbaum R.A., Foighil D.Ó., Parenti L.R., Parham J.F., Paul E., Paulay G., Pérez-Emán J., Pérez-Matus A., Poe S., Pogonoski J., Rabosky D.L., Randall J.E., Reimer J.D., Robertson D.R., Rödel M.-O., Rodrigues M.T., Roopnarine P., Rüber L., Ryan M.J., Sheldon F., Shinohara G., Short A., Simison W.B., Smith-Vaniz W.F., Springer V.G., Stiassny M., Tello J.G., Thompson C.W., Trnski T., Tucker

P., Valqui T., Vecchione M., Verheyen E., Wainwright P.C., Wheeler T.A., White W.T., Will K., Williams J.T., Williams G., Wilson E.O., Winker K., Winterbottom R., Witt C.C. 2014. Specimen collection: an essential tool. Science. 344:814–815.

Salis A.T., Ali N., Karmakar S., Kaur J., Smith B.T., Raxworthy C.J. 2026. A comparative survey of historical DNA degradation from vertebrate museum specimens: Unexpected utility of fluid-preserved specimens. Mol. Ecol. Resour. 26:e70206.

Salis A.T., Watson Z., Forcellati M., Ali N., Karmakar S., Kaur J., Rabibisoa N., Raxworthy C.J., Smith B.T. 2025. Unrecognised DNA degradation in flash-frozen genetic samples in natural history collections. Mol. Ecol. Resour.:e14121.

Schindel D.E., Cook J.A. 2018. The next generation of natural history collections. PLoS Biol. 16:e2006125.

Shi B., Berijanian M., Clark C.R., Deatherage T., Jahanshahi I., Manandhar I., Mattison Q., Quesada K.C., Schofield J., Selvick S., Yourganov G., Lytle N. 2025. MENTOR-CS: Management education for navigating transdisciplinary online research in computer science. 2025 IEEE Digital Education and MOOCS Conference (DEMOcon).:1–6.

Sigwart J.D., Schleuning M., Brandt A., Pfenninger M., Saeedi H., Borsch T., Häffner E., Lücking R., Güntsch A., Trischler H., Töpfer T., Wesche K., Collectomics Consortium. 2025. Collectomics – towards a new framework to integrate museum collections to address global challenges. Natural History Collections and Museomics. 2:1–20.

Sterner B., Elliott S., Gilbert E.E., Franz N.M. 2023. Unified and pluralistic ideals for data sharing and reuse in biodiversity. Database (Oxford). 2023:baad048.

Thomson S.A., Pyle R.L., Ahyong S.T., Alonso-Zarazaga M., Ammirati J., Araya J.F., Ascher J.S., Audisio T.L., Azevedo-Santos V.M., Bailly N., Baker W.J., Balke M., Barclay M.V.L., Barrett R.L., Benine R.C., Bickerstaff J.R.M., Bouchard P., Bour R., Bourgoin T., Boyko C.B., Breure A.S.H., Brothers D.J., Byng J.W., Campbell D., Ceríaco L.M.P., Cernák I., Cerretti P., Chang C.-H., Cho S., Copus J.M., Costello M.J., Cseh A., Csuzdi C., Culham A., D'Elía G., d'Udekem d'Acoz C., Daneliya M.E., Dekker R., Dickinson E.C., Dickinson T.A., van Dijk P.P., Dijkstra K.-D.B., Dima B., Dmitriev D.A., Duistermaat L., Dumbacher J.P., Eiserhardt W.L., Ekrem T., Evenhuis N.L., Faille A., Fernández-Triana J.L., Fiesler E., Fishbein M., Fordham B.G., Freitas A.V.L., Friol N.R., Fritz U., Frøslev T., Funk V.A., Gaimari S.D., Garbino G.S.T., Garraffoni A.R.S., Geml J., Gill A.C., Gray A., Grazziotin F.G., Greenslade P., Gutiérrez E.E., Harvey M.S., Hazevoet C.J., He K., He X., Helfer S., Helgen K.M., van Heteren A.H., Hita Garcia F., Holstein N., Horváth M.K., Hovenkamp P.H., Hwang W.S., Hyvönen J., Islam M.B., Iverson J.B., Ivie M.A., Jaafar Z., Jackson M.D., Jayat J.P., Johnson N.F., Kaiser H., Klitgård B.B., Knapp D.G., Kojima J.-I., Kõljalg U., Kontschán J., Krell F.-T., Krisai-Greilhuber I., Kullander S., Latella L., Lattke J.E., Lencioni V., Lewis G.P., Lhano M.G., Lujan N.K., Luksenburg J.A., Mariaux J., Marinho-Filho J., Marshall C.J., Mate J.F., McDonough M.M., Michel E., Miranda V.F.O., Mitroiu M.-D., Molinari J., Monks S., Moore A.J., Moratelli R., Murányi D., Nakano T., Nikolaeva S., Noyes J., Ohl M., Oleas N.H., Orrell T., Páll-Gergely B., Pape T., Papp V.,

Parenti L.R., Patterson D., Pavlinov I.Y., Pine R.H., Poczai P., Prado J., Prathapan D., Rabeler R.K., Randall J.E., Rheindt F.E., Rhodin A.G.J., Rodríguez S.M., Rogers D.C., Roque F. de O., Rowe K.C., Ruedas L.A., Salazar-Bravo J., Salvador R.B., Sangster G., Sarmiento C.E., Schigel D.S., Schmidt S., Schueler F.W., Segers H., Snow N., Souza-Dias P.G.B., Stals R., Stenroos S., Stone R.D., Sturm C.F., Štys P., Teta P., Thomas D.C., Timm R.M., Tindall B.J., Todd J.A., Triebel D., Valdecasas A.G., Vizzini A., Vorontsova M.S., de Vos J.M., Wagner P., Watling L., Weakley A., Welter-Schultes F., Whitmore D., Wilding N., Will K., Williams J., Wilson K., Winston J.E., Wüster W., Yanega D., Yeates D.K., Zaher H., Zhang G., Zhang Z.-Q., Zhou H.-Z. 2018. Taxonomy based on science is necessary for global conservation. PLoS Biol. 16:e2005075.

Webster M.S. 2017. The Extended Specimen. Philadelphia, PA: CRC Press.

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. PLoS One. 7:e29715.

Wilkinson M.D., Dumontier M., Aalbersberg I.J.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., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C.T., Finkers R., Gonzalez-Beltran A., Gray A.J.G., Groth P., Goble C., Grethe J.S., Heringa J., 't Hoen P.A.C., Hooft R., Kuhn T., Kok R., Kok J., Lusher S.J., Martone M.E., Mons A., Packer A.L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S.-A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M.A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 3:160018.

Zizka A., Silvestro D., Andermann T., Azevedo J., Duarte Ritter C., Edler D., Farooq H., Herdean A., Ariza M., Scharn R., Svantesson S., Wengström N., Zizka V., Antonelli A. 2019. CoordinateCleaner : Standardized cleaning of occurrence records from biological collection databases. Methods Ecol. Evol. 10:744–751.

Zwickl D.J., Hillis D.M. 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51:588–598.