Biodiversity Informatics, 20, 2026, pp. 118-131

ADAPTING DIGITAL HUMANITIES TOOLS FOR BIODIVERSITY SCIENCE: A CASE STUDY IN FIELD NOTEBOOK DIGITIZATION AND THE EXTENDED SPECIMEN NETWORK

Andrea Weeks*[1], Jason H. Best[2], Wendy Mann[1], Mare Nazaire[3], Ana Niño[2], Matthew Sheik[4], Krishna B. Shenoy[2], Alyssa T. Fahringer[1]

*[1]*George Mason University,* [2]*Botanical Research Institute of Texas,* [3]*California Botanic Garden,* [4] *Denver Botanic Gardens*

Abstract: Field notebooks represent invaluable primary sources of biodiversity information, yet their rate of digitization has lagged behind that of natural history specimens. This disparity stems from multiple challenges, including the perception of notebooks as personal rather than institutional records, their complex and idiosyncratic content, and the lack of standardized protocols for creating discoverable linkages between field notebooks and their associated specimens—a key component of the Extended Specimen Network concept. This paper presents a case study demonstrating an efficient, specimen-first protocol for digitizing field notebooks and establishing stable connections to the natural history specimens generated during collection events. The method was developed during curation of the salvaged Lord Fairfax Community College herbarium, which included 16 hand-written field notebooks indexed to its specimens. We adapted Omeka S, an open-source content management system designed for cultural heritage institutions, to host, view, and interpret the field notebooks within a publicly accessible digital framework. Our protocol involves three key components: 1) digitizing field notebooks and creating persistent linkages via Archival Resource Keys between notebook entries and indexed herbarium specimen records within the Symbiota database platform; 2) enabling reciprocal discovery between the associated digitized datasets for annotation and interpretation; and 3) describing these associations using controlled vocabulary, making them discoverable within Symbiota. This approach demonstrates how digital humanities tools can advance biodiversity science, particularly benefiting institutions seeking to leverage field notebook data to improve the accuracy and precision of their specimen records and to address biodiversity knowledge shortfalls. While most appropriate for field notebooks with predominantly indexed specimen collection events, our method maintains accessibility for future semantic annotation of illustrations and textual data not directly related to individual specimens.

Keywords: extended specimen, field notebooks, herbaria, Omeka S, Symbiota

Introduction

Field notebooks are well-regarded primary sources of information about biodiversity. Created by naturalists, explorers and collectors of natural history specimens, field notebooks record first-hand observations of places and organisms in time that, in some cases, are tied to specimen collection events. These types of documents, often hand-written and illustrated, have many different names, such as accession logs, collection notes, field books, journals or ledgers, but all typically have idiosyncratic notation styles unique to their authors (Lockshin and Bennett 2018). Research scientists and museum curators have long referred to field notebooks to detect and correct errors in specimen collection labels and improve their interpretation of the specimen and, by extension, the species it represents (e.g., Palandačić et al. 2024, Peterson et al. 2004, Weeks 2024, Woller and Hill 2015). Observations not tied to individual specimens but still concerning species' habitat, such as climate or water conditions, or apparent biology, such as community-level abundance or reproductive behavior, have equal value for addressing shortfalls in our knowledge about biodiversity (Astudillo-Clavijo et al.

* Correspondence: aweeks3@gmu.edu

2024, Reznick et al. 1994). Indeed, early discussions about the ways in which widescale natural history specimen digitization could advance biodiversity science incorporated specific recommendations for including field notebook data (Graham et al. 2004, Tingley and Beissinger 2009). More recently, the first articulation of the extended specimen concept (Webster 2017) included field notes as one of the ancillary datasets for natural history specimens. This concept centers the importance of physical natural history specimens for verifying, organizing and discovering the numerous types of data associated with or later derived from these specimens.

Yet, despite advances in digitization and substantial collaborative efforts among natural history museums and libraries in the past two decades, digital access to field notebook data has lagged behind that for natural history specimens. For comparison to date, two prominent field book digitization campaigns, the BHL Field Notes Project[1] and the Smithsonian Field Book Project[2,3], have collectively transcribed and digitized 6,142 field notebook titles comprising >560,000 pages, whereas the largest biodiversity data aggregator, the Global Biodiversity Information Facility (GBIF)[4], hosts >120,000 datasets containing >284 million species occurrence records from preserved and fossil natural history specimens. The reason for this imbalance, beyond lack of resources, may be due to the attributes of field notebooks that natural history specimens do not share: the longstanding perception of field notebooks as personal possessions rather than as museum records (Jones 2017) and the complexity of accurately cataloging the many scientific and cultural-historical data types within them (Kearney 2015, Nakasone and Sheffield 2013, Rinaldo et al. 2023, Sheffield and Nakasone 2011, Stork et al. 2019b). Field notebooks are often discarded or become disassociated with the specimens they describe and are therefore challenging to relocate or may be housed in institutions that do not have the resources to completely digitize them. Surmounting these barriers for the global store of historical and contemporary field notebooks will be essential for achieving the Extended Specimen Network (BCN 2019, Hardisty et al. 2022, Kapun et al. 2025), which envisions a findable, accessible, interoperable and reusable digital environment (FAIR; Wilkinson et al. 2016) that durably interconnects natural history specimen records with their constellation of ancillary and derivative datasets.

Incorporating field notebook data in this next stage of digitization for natural history collections ("Digitization 2.0"; Hedrick et al. 2020) presents a significant challenge and opportunity to the biodiversity informatics community. Some larger museums with customized natural history collection databases associate field notebook media records with their specimen records, such as MCZBase[5], and assign unique identifiers to these digital resources (Astudillo-Clavijo et al. 2024), although these associations may not be accessible via biodiversity data aggregators. While Darwin Core vocabulary (Wieczorek et al. 2012, Darwin Core Maintenance Group 2025) does contain accepted terms that allow occurrence records to be amended with text strings associated with field notes (dwc:fieldNotes) or undefined values (dwciri:fieldNotes), such as universal resource locators (URLs), semantic vocabulary for describing complex resource relationships is not necessarily incorporated into collections' databases or shared to biodiversity data aggregators.

Where does the current state of digitization leave natural history collections with field notebooks that want to leverage this rich source of biodiversity information within a digital framework, especially those of small institutions or institutions residing in the Global South that provide unique and outsized contributions to global biodiversity knowledge (Delves et al. 2024, Marsico et al. 2020)? There are no widely accepted protocols for digitizing field notebooks as there are for natural history specimens (e.g., Nelson et al. 2015), and none, to our knowledge, that describe how to create the semantically detailed resource relationship between field notebook records and their specimen records as described by the calls for the Extended Specimen Network. We confronted this issue when the herbarium of George Mason University salvaged specimens from the Lord Fairfax Community College herbarium (LFCC; Weeks 2020), which included 16 volumes of hand-written field notebooks whose entries were numerically indexed to thousands of individual specimens. These notebooks were created by the founding curator of LFCC, Professor Robert Charles Simpson (Simpson 2026), and comprise >4000 pages of loose-leaf data-collection sheets and free-form text entries and illustrations. As part of a U.S. National Science Foundation funded project[6] to curate the salvaged specimens of LFCC and make their metadata and high-resolution images publicly accessible as digital research infrastructure, we adapted open-source software tools from the digital humanities to create an efficient *specimen-first* protocol that 1) digitizes field notebooks and creates persistent URL linkages between their entries and their indexed herbarium specimen records, 2) allows reciprocal discovery between the associated digi-

[1] https://www.biodiversitylibrary.org/collection/FieldNotesProject.
[2] https://www.biodiversitylibrary.org/collection/SIFieldbooks.
[3] https://siarchives.si.edu/about/field-book-project.
[4] https://www.gbif.org/.
[5] https://mczbase.mcz.harvard.edu/.
[6] https://www.nsf.gov/awardsearch/show-award/?AWD_ID=2022918.

Table 1. Omeka resource template maps Symbiota, Omeka S and corresponding Dublin or Darwin Core vocabulary.

Original label	Alternate label	Vocabulary	Term
Identifier		Dublin Core	dcterms:identifier
Title	Barcode	Dublin Core	dcterms:title
Catalog Number		Darwin Core	dwc:catalogNumber
Other Catalog Numbers	LFCC Accession Number	Darwin Core	dwc:otherCatalogNumbers
Family		Darwin Core	dwc:family
Scientific Name		Darwin Core	dwc:scientificName
Scientific Name Authorship	Author	Darwin Core	dwc:scientificNameAuthorship
Recorded By	Collector	Darwin Core	dwc:recordedBy
Record Number	Number	Darwin Core	dwc:recordNumber
Event Date		Darwin Core	dwc:eventDate
Identified By		Darwin Core	dwc:identifiedBy
Date Identified		Darwin Core	dwc:dateIdentified
Country		Darwin Core	dwc:country
State Province	State/Province	Darwin Core	dwc:stateProvince
County		Darwin Core	dwc:county
Locality		Darwin Core	dwc:locality
Minimum Elevation In Meters	Elev. Min. (m)	Darwin Core	dwc:minimumElevationInMeters
Maximum Elevation In Meters	Elev. Min. (m)	Darwin Core	dwc:maximumElevationInMeters
Verbatim Elevation	Verbatim Elev.	Darwin Core	dwc:verbatimElevation
Habitat		Darwin Core	dwc:habitat
Description		Dublin Core	dcterms:description
Basis of Record		Darwin Core	dwc:basisOfRecord

tized datasets and their subsequent annotation and 3) describes these associations using a controlled vocabulary within a publicly accessible digital portal for the herbarium specimen data.

We provide here a case study about our methods, results and a discussion of limitations and benefits of a streamlined approach to making field notebook data digitally accessible as a tool for curators and discoverable by researchers for further use and interpretation. Our approach includes adapting the Omeka S software platform to host, view and interpret field notebook images and textual data. Omeka[7] is a content management system originally designed for galleries, libraries, archives, and museums to create online collections and narrative digital exhibits. Since its initial release in 2008, it has been downloaded over half a million times and is used throughout the world by cultural heritage institutions, scholars, educators, and enthusiasts. Development of Omeka S, a version of the software made for institutions wanting to manage multiple websites from one administrative interface and share content amongst them, began in 2012. While the software is free and open source, users must have server space to host the installation. Most published use cases of Omeka S incorporate digital data from the humanities, some that explore the history of science and ethnobiology include interpretation of natural history illustrations and objects[8,9]. Insofar as natural history collections contribute to research in the humanities (Jacobs and Griebler 2025), we believe our adaptation of Omeka S for field notebooks illustrates how the digital humanities can advance biodiversity science.

[7] https://omeka.org/.

[8] https://exhibit.utas.edu.au/s/olive-pink/page/welcome.

[9] https://museumcollab.anthro.lsa.umich.edu/s/Anishinaabe/.

Methods

Herbarium specimen digitization

Digitization of the LFCC herbarium specimens followed the Nelson et al. (2015) protocol and used Symbiota database software (Gries et al. 2014) hosted live on the publicly accessible SouthEast Regional Network of Expertise and Collections (SERNEC) portal[10]. During manual transcription of the herbarium specimen labels, extra attention was taken in distinguishing personal collection numbers from the LFCC accession numbers, both of which are unique to each specimen. These values were recorded within database fields corresponding to the Darwin Core terms, dwc:recordNumber and dwc:OtherAccessionNumber, respectively. Both personal collection and LFCC accession numbers were used to link the Symbiota occurrence record to the indexed notebook entries within Omeka S. After the notebooks were imaged and parsed within Omeka S (see following subsections), the persistent URLs of the notebook pages displaying the handwritten entries for individual herbarium specimens were imported to the Symbiota records for the herbarium specimens. Within Symbiota, the URLs were classified as data associations ("Associations") that referenced the specimen occurrence ("isReferencedBy") and were tagged as notebook records ("fieldNotes"). The completed digitized herbarium dataset was shared to iDigBio[11] and GBIF (LFCC 2026).

Notebook digitization

Initial digitization of the LFCC notebooks was completed by Backstage Library Works (Bethlehem, Pennsylvania, USA), a vendor specializing in the digitization, description and management of archival collections. The pages from the notebook volumes were imaged as separate TIFF files, and the personal collection number or LFCC accession number listed on each page, depending on the notebook, was manually keyed into an indexed spreadsheet as well as embedded in the TIFF headers. PDF files of the notebook volumes and JPG derivatives for online display were created from the original TIFF files. A copy of the digital media files produced by the vendor was installed on a George Mason University Libraries server for later integration with the Omeka S platform. After digitization, the physical notebooks were accessioned into the Special Collections Research Center of the George Mason University Libraries (Lord Fairfax Community College Herbarium notebooks collection, 1973–1992), and their library catalog record was later amended with a URL pointing to their digital surrogates on the Omeka S platform.

Omeka S installation and customization

The George Mason University Libraries installed a copy of Omeka S for the LFCC project on one of its servers. Within this installation, we customized the content and software functionalities to generate the interactive online site that enables site visitors to explore the digitized notebooks and their interpretative content gleaned from LFCC herbarium records (Fahringer et al. 2026). The public-facing site uses the Center Row design theme and includes "Welcome" and "About" pages, both of which utilize HTML content blocks, as well as "Browse" and "Search" pages for the notebooks. The search page is a custom URL that is linked to the advanced search functionality. Omeka S installations include Dublin Core (Dublin Core Metadata Initiative, 2020), an all-purpose metadata vocabulary, and this vocabulary was used to describe some aspects of the notebook and specimens. However, to accurately describe and manage in Omeka S the specialized natural history specimen metadata, we imported Darwin Core terms to Omeka S using the "import new vocabulary" function and the following specifications: Label: Darwin Core, Namespace URI: http://rs.tdwg.org/dwc/terms/, Namespace prefix: dwc, Import type: URL, File URL: https://rs.tdwg.org/dump/terms.ttl.

Two item sets were created within the Omeka S installation to manage the digitized notebook files: the Robert Simpson Notebooks, comprising the 16 field notebooks that were curated by the founding curator of LFCC, uploaded in a manner that allows visitors to view and explore them in their entirety as PDF documents, and the LFCC Specimen Collection and Accession Records, which include the extant LFCC herbarium specimen records that are indexed to separate notebook pages. We created a resource template (Table 1), which ensured consistency when describing items and made browsing and searching for items on the public-facing site more efficient. From the administrative dashboard, we used the "add new resource template" feature to provide a label for the template and added fields by searching through available properties. Lastly, we installed the ARK module to create Archival Resource Keys (ARKs), a type of persistent URL identifier, for each item within Omeka S, including the 16 notebooks and every item comprising an LFCC accession number or personal collection number with its corresponding notebook page. This module was obtained using the Name Assigning Authority Number (NAAN) request form with the ARK Alliance[12], which is required if one's organization does not already have a NAAN. The ARK module also required the installation of the Common module, which provides essential functions that are shared by

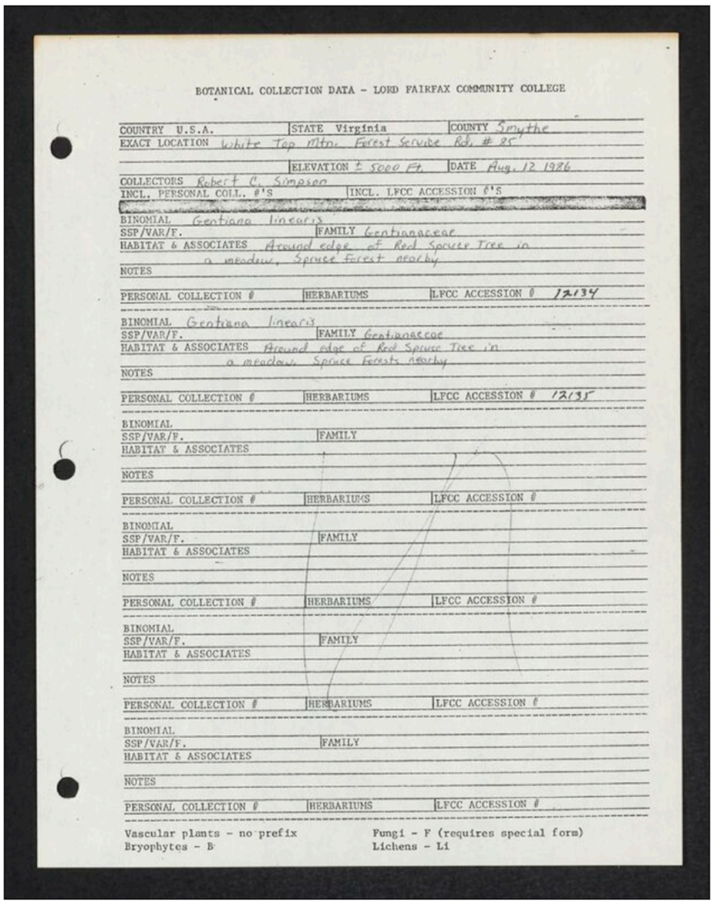
[10] https://sernecportal.org/portal/collections/misc/collprofiles.php?collid=441.
[11] https://www.idigbio.org/.
[12] https://arks.org/.

LORD FAIRFAX COMMUNITY COLLEGE HERBARIUM (LFCC) SPECIMEN COLLECTION AND ACCESSION RECORDS

Welcome | About | Browse specimens | Browse notebooks | Search

Notebook16_097.jpg

Item


BOTANICAL COLLECTION DATA – LORD FAIRFAX COMMUNITY COLLEGE

COUNTRY U.S.A. | STATE Virginia | COUNTY Smythe
EXACT LOCATION White Top Mtn. Forest Service Rd. # 85
ELEVATION ± 5000 Ft. | DATE Aug. 12 1986
COLLECTORS Robert C. Simpson
INCL. PERSONAL COLL. #'S | INCL. LFCC ACCESSION #'S
[redacted]
BINOMIAL Gentiana linearis
SSP/VAR/F. | FAMILY Gentianaceae
HABITAT & ASSOCIATES Around edge of Red Spruce Tree in a meadow, Spruce Forest nearby
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION # 12134

BINOMIAL Gentiana linearis
SSP/VAR/F. | FAMILY Gentianaceae
HABITAT & ASSOCIATES Around edge of Red Spruce Tree in a meadow. Spruce Forests nearby
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION # 12135

BINOMIAL
SSP/VAR/F. | FAMILY
HABITAT & ASSOCIATES
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION #

BINOMIAL
SSP/VAR/F. | FAMILY
HABITAT & ASSOCIATES
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION #

BINOMIAL
SSP/VAR/F. | FAMILY
HABITAT & ASSOCIATES
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION #

BINOMIAL
SSP/VAR/F. | FAMILY
HABITAT & ASSOCIATES
NOTES
PERSONAL COLLECTION # | HERBARIUMS | LFCC ACCESSION #

Vascular plants – no prefix   Fungi – F (requires special form)
Bryophytes – B   Lichens – Li

Identifier	https://sernecportal.org/portal/collections/individual/index.php?occid=22920719 ark:/61221/25761
Barcode	Notebook16_097.jpg
Catalog Number	LFCC-0002011
LFCC Accession Number	12134
Family	Gentianaceae
Scientific Name	Gentiana linearis

Figure 1. Example of the Omeka S record of the field notebook page that describes the collecting event for LFCC Accession #12134, https://masonlibraries.gmu.edu/lfcc/s/lfcc/item/25761 . Bottom of page image is cropped for display and excludes additional information fields about the item (see Table 1).

Details | Map | Comments | Linked Resources | Edit History

LFCC Lord Fairfax Community College Herbarium (LFCC)

Catalog #: LFCC-0002011
Occurrence ID: bfdf6c0f-ac1f-404d-a170-3a61a09b0757
Secondary Catalog #: 12134
Taxon: *Gentiana linearis* Froel.
Family: Gentianaceae
Determiner: G.P. Fleming, T.F. Wieboldt, and J.F. Townsend
Date Determined: 11/16/19
⊞ Show Determination History
Collector: Robert C. Simpson
Number: s.n.
Date: 1986-08-12
Locality: United States, Virginia, Smyth, White Top Mtn., Forest Service Rd. #85
Latitude/Longitude: 36.641341 -81.60508 +-1457m. WGS84
Verbatim Coordinates: 17S 445910.26E 4055256.82N
Elevation: 1524 meters Verbatim Elevation: +/-5000 ft.
Habitat: Around edge of Red Spruce tree in a meadow. Spruce forest nearby.

Associations
isReferencedBy (fieldNotes):
https://n2t.net/ark:/61221/25761

Specimen Media

Open Medium File
Open Large File

Figure 2. Example of the Symbiota record for the herbarium specimen LFCC Accession #12134, https://sernecportal.org/portal/collections/individual/index.php?occid=22920719. Note the ARK URL within the Associations call-out box that links to the field notebook entry for this collection within the Omeka S platform and is described using controlled vocabulary.

different modules. Once the ARK module was installed and configured, an ARK was issued to each new item at the time of its creation and stored within the Dublin Core "Identifier" field (i.e., dcterms:identifier).

Adding metadata and image files to Omeka S

We relied exclusively on the CSV import function when creating items in Omeka S. Metadata for each herbarium specimen, including the Symbiota URL for the public record of the herbarium specimen and its digital image, was extracted from the LFCC Symbiota database as a spreadsheet with column headers matching the terms identified in the LFCC resource template. This enabled correctly matching herbarium specimens to their corresponding notebook page entry, either via personal collection number or LFCC accession number. From the administrative dashboard, we imported the file using the default options and mapped it to Omeka S terms using the resource template (Table 1). The CSV import module can be used to append, revise, update, or replace metadata. We attached notebook image files to items using the File Sideload module in conjunction with CSV import and configured the module to include the absolute path to the directory where the notebook image files were stored.

Results

Integrating the digitized records for LFCC notebooks and herbarium specimens allowed detailed comparison of the datasets and revealed unexpected differences between them. The 16 LFCC notebooks comprise 4,053 pages of hand-written specimen collection and accessioning events that include 12,090 LFCC accession numbers and 1,216 personal collection numbers, all of which reference physical herbarium specimens created for and/or deposited at LFCC. Of these 13,306 events, 10,609 (86%) are matched definitively to one or more (duplicate) LFCC herbarium specimens within Omeka S. The Omeka S item page for each of these specimen collection events contains imported herbarium specimen metadata and the Symbiota URL for the public record of the herbarium specimen (Figure 1). The extant LFCC herbarium contains 12,857 specimens of which 11,026 herbarium specimens (86%) are referenced by the 10,609 notebook entries. The referenced herbarium specimens outnumber the matching entries due to the large number of duplicate herbarium specimens in the collection. The Symbiota database records of these 11,026 specimens include the ARK of its Omeka S item page, whose association within the Symbiota database is referenced using the controlled vocabulary "isReferencedBy (fieldNotes)" (Figure 2). The remaining 14% of the extant LFCC herbarium specimens either lack sufficient label information to provide a definitive connection to a notebook entry or were never logged into the notebooks. Conservatively estimated, 449 LFCC herbarium specimens referenced by the notebooks cannot not be accounted for; it is possible some or all of these were lost or destroyed prior to the salvage of the herbarium collection in 2019. We are aware that at least one page from the notebooks is missing (i.e., Notebook 5, page(s) containing LFCC Accession numbers 11–14) and that several other herbarium specimens with LFCC accession numbers lack corresponding entries in the notebooks, so there remains the possibility that other notebook records have been lost as well. Reference to individual notebook pages and their entries during later curation of the LFCC specimen records was streamlined because of the digital linkages between the datasets. The range of biodiversity knowledge uncovered within the notebooks during this process is reported in Table 2 and discussed in the next section.

Of the 11,026 LFCC extant herbarium specimens referenced by collection events in the notebooks, the majority (53%, 5852 specimens) were collected by Robert Simpson alone or in association with others, as expected. The remaining 5,174 LFCC specimens referenced in the notebooks were collected by other people, whose names have not been disambiguated completely to provide a precise count. However, 528 different name strings are listed as the primary collector(s) for these specimens. These name strings include multiple names or spelling variants of the same name so are only an approximation of the number of different collectors. The number of specimens collected per name string ranges from one to 242. Conservatively estimated, these other collectors of LFCC specimens tally in the hundreds yet on an individual basis contributed very few specimens to the herbarium, as compared to Robert Simpson. The identity of these other collectors and their role in authoring the notebook entries are explored further in the discussion.

Discussion

In contrast to field notebook digitization techniques that create a fully transcribed document or combine targeted transcription with extensive semantic annotation of text and image data on each page (Kearney 2015, Stork et al. 2019b), the protocol outlined here transcribes very little of the original data within the field notebooks. Instead, transcription is limited to the collection numbers or accession numbers associated with specimen collection events on the pages, which are appended as metadata to the images of the notebook pages. On the Omeka S platform, the textual information that interprets field notebook entries and the images of their corresponding notebook pages are

Table 2. Examples of data within LFCC notebooks that address biodiversity knowledge shortfalls sensu Hortal et al. (2015) and improve research capacity.

New datum discovered in notebooks		Example of its effect improving biodiversity research [shortfall addressed]	Notebook entry page(s) (URL) example	Herbarium specimen referenced by notebook entry (URL) example
1	date of collection	expands capacity for phenological research [Raunkiaeran]	https://n2t.net/ark:/61221/19717 (specimen entry) https://n2t.net/ark:/61221/19701 (collection event entry)	https://sernecportal.org/portal/collections/individual/index.php?occid=24463821
2	detailed locality	permits retrospective assignment of geocoordinates for species-distribution modeling [Wallacean]	https://n2t.net/ark:/61221/3114	https://sernecportal.org/portal/collections/individual/index.php?occid=28202402
3	plant abundance	allows longitudinal population assessment of species [Prestonian]	https://n2t.net/ark:/61221/1440	https://sernecportal.org/portal/collections/individual/index.php?occid=22899039
4	timing of floral scent release	advances understanding of reproductive strategies [Raunkiaeran]	https://n2t.net/ark:/61221/1287	https://sernecportal.org/portal/collections/individual/index.php?occid=22899379
5	habitat and substrate	contributes information about adaptations to abiotic conditions [Hutchinsonian]	https://n2t.net/ark:/61221/5309	https://sernecportal.org/portal/collections/individual/index.php?occid=27601961
6	associated taxa	contributes information about potential biotic interactions [Eltonian]	https://n2t.net/ark:/61221/25800	https://sernecportal.org/portal/collections/individual/index.php?occid=22878774
7	datum that reveals an error reported on herbarium specimen	increases the accuracy and precision of biodiversity data	https://n2t.net/ark:/61221/16982	https://sernecportal.org/portal/collections/individual/index.php?occid=24026747
8	an original anatomical drawing justifying species determination	provides a visual guide for accurate species identification	https://n2t.net/ark:/61221/1275 (specimen entry) https://n2t.net/ark:/61221/1272 (drawing)	https://sernecportal.org/portal/collections/individual/index.php?occid=24131819
9	the other institutions housing duplicate specimens	assists finding specimens from the same individual for research	https://n2t.net/ark:/61221/809	https://sernecportal.org/portal/collections/individual/index.php?occid=23889635

10	collection events of a rare species that reference missing herbarium specimens	changes species occurrence status from waif individual to possibly established population [Linnean]	https://n2t.net/ark:/61221/25762	https://sernecportal.org/portal/collections/individual/index.php?occid=31536346
11	an original observation about habitat and growth form differences between taxonomic varieties	adds insight regarding divergent traits of putatively closely related taxa [Darwinian]	https://n2t.net/ark:/61221/1684	n/a
12	collector's identity	connects the collector's biography to the specimen as an additional source of relevant data	https://n2t.net/ark:/61221/20349	https://sernecportal.org/portal/collections/individual/index.php?occid=31545585
13	a species determination not reflected by specimen annotation labels	informs the history of professional scientific networks in biodiversity exploration	https://n2t.net/ark:/61221/809	https://sernecportal.org/portal/collections/individual/index.php?occid=23889635
14	description of specimen collection methods and field data notation protocol	reveals the motivations and biases of the specimen collectors	https://n2t.net/ark:/61221/28217 (first page) https://n2t.net/ark:/61221/28219 (second page)	n/a

imported from herbarium specimen records, whose label data were previously transcribed into Symbiota database fields. This lean approach, which we refer to as a specimen-first or specimen-centric approach, prioritizes rapidly and durably interconnecting specimen records and their field notebook entries for later inspection and interpretation. Curators from within their Symbiota specimen record database may easily reference notebook entries to update specimen records or generate annotation labels for the physical specimens citing the notebook entries. Public viewers of the specimen records within the Symbiota specimen data portal may also read these notebook entries and leave comments to the curators about their interpretation. Visitors to the Omeka S site may search the notebook entries using any of the Darwin Core fields that were imported or the image metadata tags, browse the entirety of the notebooks, and follow linkages back to the Symbiota specimen record database. Although Omeka S can accommodate public comment dialog boxes for items, which could facilitate greater community-level annotation of these resources in line with the recommendations of Astudillo-Clavijo et al. (2024), the LFCC project limits public comment to the Symbiota site at this time.

The specimen-first approach to digitizing the LFCC field notebooks created an important curatorial tool. Unexpected data about the LFCC herbarium and its individual specimens were routinely discovered within their notebook entries that emended or expanded the metadata available from the herbarium specimen labels alone. In some cases (Table 2, examples 1–6), these new data allowed curatorial updates to the records for individual specimens that collectively improved their capacity to address biodiversity knowledge shortfalls. The digital records of the specimens were amended to resolve missing dates of collection, precise locality descriptions, plant abundances, floral scent descriptions, habitat and substrate type, and associated taxa, which inform multiple dimensions of a species' biology. Such data, for example, can be used to refine understanding of a species' phenology, train its species-distribution model, track its population size through time, investigate its reproductive biology as well evaluate its abiotic and biotic ecological interactions, respectively. In other cases (Table 2, example 7), notebook entries revealed that their associated herbarium specimens contained incorrect label information, so we emended their original metadata records with the corrections provided by the notebooks.

The LFCC field notebooks also revealed new information that exceeded the metadata typically recorded or shared publicly for individual herbarium specimens yet are essential for advancing biodiversity research. These

included anatomical drawings justifying species determinations, the names of other institutions that received duplicate specimens, collection events of rare species that reference apparently missing herbarium specimens, and original observations about habitat and growth form differences between taxonomic varieties (Table 2, examples 8–11). For example, the Virginia state record of *Gentiana linearis* Froel. was first discovered in the LFCC collection and was assumed to be a waif occurrence (Figure 1 and 2, LFCC Accession #12234; Table 2, example 9). However, on inspection of the notebook records, it was revealed that a second individual had been collected concurrently (LFCC Accession #12235), which increased the likelihood that these occurrences documented an established population. Later, the specimen of this second individual was discovered among sheets modified with obscured labels for teaching purposes.

The field notebook data that revealed the identities, motivations and biases of the people who created the specimens and ultimately authored the biodiversity information (Table 2, examples 12–14) added important context for the collection. They are also a type of cultural-historical information common within field notebooks that are relevant to the humanities. Some notebook entries revealed the full name of the collector, which allowed the specimen to be added to the known exsiccatae for that botanist. For example, the collector of LFCC Accession #9437, Marlin Krouse (Krouse 2026; Table 2, example 12), frequently accompanied botanist Lena Artz (Artz 2026; Weeks 2024) to sites in the Massanutten Mountains of Virginia. From this biographical information, one can pinpoint the Laurel Run locality for his #9437 collection in these mountains. Other entries note professional networks for individual specimen determinations that are not listed on annotation labels (Table 2, example 13) and that shed light on the people that Professor Simpson relied on early in his career for confirming plant identifications.

A particularly critical discovery was a two-page preface to a notebook describing specimen collection methods and the field data notation protocol (Table 2, example 14) that provides information about the genesis of the herbarium and the authorship of its field notebooks. For instance, these pages explain why virtually no LFCC herbarium specimens record plant height, as this trait was not part of the required field notation protocol. The pages also reveal that the herbarium was, in large part, assembled by undergraduate students instructed by Professor Simpson. Inspection of the >4000 pages of the LFCC field notebooks reveals many different handwriting styles, which indicates that the observations were recorded by other individuals in addition to Professor Simpson. In some cases, as inferred by matching handwriting samples, individuals who wrote the field notebook entries also wrote the labels for herbarium specimens created during collection events, which suggests these individuals were the collectors of the specimens and perhaps also the students of Professor Simpson. As described in the results, the small number of specimens contributed by each collector name string is consistent with that required for undergraduate student projects. In other cases, individuals acted as scribes for others, perhaps recopying notes as a curatorial assistant or as part of a group project by recording their student partners' collection events. The precise number of individuals contributing to the notebook entries was not determined, but analysis of the collector name strings indicates they may tally in the hundreds. Thus, the LFCC field notebooks are unusual from typical field notebooks in having multiple authors, with the caveat that most or all of these individuals were under the direction of Professor Simpson. His students, as residents of the largely rural landscapes of northwestern Virginia in the 1970s and 1980s, had collectively wide access to private and public land on which to collect specimens and contributed to the density of herbarium specimens from the counties surrounding the college (Weeks et al. 2025). Such features of this collection would have remained unrecorded and unknown but for these two notebook pages.

Herbarium record updates in the Symbiota database are not automatically transferred to the Omeka S LFCC website after their initial import during set-up. Instead, updates must be manually imported, and the lag time between imports may result in temporarily conflicting information on the Omeka S platform. For instance, ongoing nomenclatural changes for the LFCC herbarium specimens are a common source of discrepancy between the two digital resources (see Table 2, example 3). This particular outcome is not unique to the Omeka S platform, as taxonomic name discrepancies are a well-known challenge for interpreting biodiversity datasets and have established protocols for detecting and resolving conflicts (Grenié et al. 2023). Because each field notebook entry item in Omeka S is linked to its herbarium specimen record in Symbiota, the most up-to-date nomenclature for that specimen is always accessible, if not part of the textual interpretation for the notebook entry itself.

The specimen-first approach to digitizing field notebooks described here does achieve part of the intent of the Extended Specimen Network, in that it creates findable and accessible connections between specimen records and their ancillary extensions to improve infrastructure for research. The architecture of the Symbiota portal is customized to allow users to search for specimens by association

type, which makes the persistent URL to notebook entries for specimens discoverable by users of the site. Reciprocally, the Omeka S portal leads visitors from notebook entries to specimen records. While these associations are not yet ingested by global biodiversity aggregators, their effects on improving the accuracy and precision of the specimen data shared to these publicly accessible sites are substantial, as detailed in the sections above. One stated goal of the Extended Specimen Network is to accelerate discovery across datasets to address biodiversity knowledge shortfalls, such as those outlined by Hortal et al. (2015). Using their framework (Table 2), we have shown how field notebook data can fill the knowledge gaps of all seven categories identified by these authors, which are essential for advancing macroecological and evolutionary research.

Some LFCC field notebook data that are not tied to herbarium specimen collection events are not surfaced immediately using our method as they would be if their textual and image data had been fully transcribed and semantically annotated (e.g. Table 2, example 11 and 14). However, future use of optical character recognition or artificial intelligence to transcribe and semantically annotate text and images is not precluded by our method; the archived high-resolution image files of the notebooks remain publicly accessible for these types of future analyses. Given the widely varying format and content of field notebooks, a variety of digitization protocols will be needed to capture their information effectively and efficiently. The specimen-first approach to digitization is more appropriate for those field notebooks with a predominance of indexed specimen collection events rather than unassociated natural-history observations or illustrations. Recent methods for semantically annotating text in digitized field notebooks (Stork et al. 2019a) and training image recognition software to produce taxonomic classification of illustrations in them (Stork et al. 2021) are promising avenues for making data from notebooks not directly associated with collecting events machine readable.

Envisioning the research that could be made possible by liberating biodiversity data at scale from historical or contemporary born-digital field notebooks and managing these data within a FAIR digital environment is tantalizing. This type of environment could empower taxonomists to query the global store of notebook records for a given taxon to test for patterns in the textual observations that might indicate the presence of segregate taxon. It also could allow ecologists to access machine-ready biotic and abiotic data records by geographic region accumulated across centuries of notetaking, which could reveal longitudinal changes in endangered ecosystems and assist plans for their restoration. Use-cases for a FAIR data environment of field notebook records are easily imagined, with the important caveat that unexpected new uses for these data are also likely to emerge in the future. Especially for field notebooks, this future is within reach given the technological tools now at our disposal, such as those provided by advances in library science, the digital humanities, and decades of natural history collections' digitization research and development. Creating the FAIR digital environment for field notebooks to usher in the next generation of biodiversity research will require community collaboration, creativity, and the human and financial resources to move this vision to reality.

Author Contributions

AW conceptualized and supervised the project, curated and analyzed data and led manuscript writing. AW, JHB, MN, AN and KBS created use-cases for field notebook data to develop resource relationships in Symbiota software. WM provided guidance on digitization methods and Omeka S software. MS assisted data curation. ATF curated data and led Omeka S implementation. All authors contributed to manuscript drafts.

Funding

This material is based upon work supported by a George Mason University Libraries Fenwick Fellowship and Virginia Native Plant Society research award to AW as well as US National Science Foundation Grants: DBI 2022918 (AW), DBI 2133561(MN) and DBI 2133562 (AN). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US National Science Foundation.

Acknowledgments

The authors wish to thank Elizabeth McMurchie, Glenn Montague, Judy and Larry Skog and W. Carl and Gerry Taylor for their assistance in salvaging the LFCC herbarium, as well as Katie Pearson and Ed Gilbert for implementing field notebook resource relationship functionality in Symbiota. This article is dedicated to Professor Robert C. Simpson (1948–2025), the founding curator of the Lord Fairfax Community College Herbarium in Middletown, Virginia.

Competing Interests

The authors have declared that no competing interests exist.

Literature Cited

Artz, L.C. 2026. Natural history specimens collected and/or identified and deposited. Zenodo. https://doi.org/10.5281/ZENODO.8034915

Astudillo-Clavijo, V., T. Mankis, and H. López-Fernández. 2024. Opening the museum's vault: Historical field records preserve reliable ecological data. Am. Nat. 203:305–322. https://doi.org/10.1086/728422

Biodiversity Collections Network (BCN). 2019. Extending U.S. biodiversity collections to promote research and education. American Institute of Biological Sciences, Washington, DC.

Darwin Core Maintenance Group. 2025. Darwin Core List of Terms. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/list/2025-07-10

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

Dublin Core Metadata Initiative. 2020. DCMI metadata terms. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Fahringer, A. T., W. Mann and A. Weeks. 2026. Lord Fairfax Community College Herbarium (LFCC) specimen collection and accession records. George Mason University. https://masonlibraries.gmu.edu/lfcc .

Lord Fairfax Community College Herbarium notebooks collection, 1973–1992. C0333, Special Collections Research Center, George Mason University Libraries. https://aspace.gmu.edu/repositories/2/resources/568

Graham, C., S. Ferrier, F. Huettman, C. Moritz, and A. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol. Evol. 19:497–503. https://doi.org/10.1016/j.tree.2004.07.006

Grenié, M., E. Berti, J. Carvajal-Quintero, G. M. L. Dädlow, A. Sagouis, and M. Winter. 2023. Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. Methods Ecol Evol 14:12–25. https://doi.org/10.1111/2041-210X.13802

Gries, C., E. Gilbert, and N. Franz. 2014. Symbiota – A virtual platform for creating voucher-based biodiversity information communities. Biodivers. Data J. 2:e1114. doi: https://doi.org/10.3897/BDJ.2.e1114

Hardisty, A. R., E. R. Ellwood, G. Nelson, B. Zimkus, J. Buschbom, W. Addink, R. K. Rabeler, J. Bates, A. Bentley, J. A. B. Fortes, S. Hansen, J. A. Macklin, A. R. Mast, J. T. Miller, A. K. Monfils, D. L. Paul, E. Wallis, and M. Webster. 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. BioScience 72:978–987. https://doi.org/10.1093/biosci/biac060

Hedrick, B. P., J. M. Heberling, E. K. Meineke, K. G. Turner, C. J. Grassa, D. S. Park, J. Kennedy, J. A. Clarke, J. A. Cook, D. C. Blackburn, S. V. Edwards, and C. C. Davis. 2020. Digitization and the future of natural history collections. BioScience 70:243–251. https://doi.org/10.1093/biosci/biz163

Hortal, J., F. De Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst. 46:523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Jacobs, H. L., and A. Griebeler. 2025. Herbarium data and herbaria as extended archives for humanities research. Botany 103:1–21. https://doi.org/10.1139/cjb-2025-0074

Jones, M. 2017. From personal to public: field books, museums, and the opening of the archives. Archives and Records 38:212–227. https://doi.org/10.1080/23257962.2016.1269645

Kapun, M., M. Schwentner, E. Haring, N. Akkari, A. Kroh, L. Kruckenhauser, A. Palandačić, and K. Vohland. 2025. Museomics, the extended specimen and collectomics – how to frame and name the diversity of information linked to specimens in natural history collections. NHCM 2:1–21. https://doi.org/10.3897/nhcm.2.161331

Kearney, N. 2015. Transcribing between the lines: crowd-sourcing historic data collection. MWA2015: Museums and the Web Asia 2015. https://mwa2015.museumsandtheweb.com/paper/transcribing-between-the-lines-crowd-sourcing-historic-data-collection/index.html

Krouse, M. 2026. Natural history specimens collected and/or identified and deposited. Zenodo. https://doi.org/10.5281/ZENODO.8080168

Lockshin, N. S., and R. W. Bennett. 2018. Smudges, snakeskins, and pins, oh my! Book and Paper Group Annual 37: 125–142. https://repository.si.edu/handle/10088/97808

Lord Fairfax Community College Herbarium (LFCC). 2026. Lord Fairfax Community College Herbarium. Occurrence dataset https://doi.org/10.15468/c2gj2t

Lord Fairfax Community College Herbarium notebooks collection, 1973–1992. C0333, Special Collections

Research Center, George Mason University Libraries. https://aspace.gmu.edu/repositories/2/resources/568

Marsico, T. D., E. R. Krimmel, J. R. Carter, E. L. Gillespie, P. D. Lowe, R. McCauley, A. B. Morris, G. Nelson, M. Smith, D. L. Soteropoulos, and A. K. Monfils. 2020. Small herbaria contribute unique biogeographic records to county, locality, and temporal scales. Am. J. Bot. 107:1577–1587. https://doi.org/10.1002/ajb2.1563

Nakasone, S., and C. Sheffield. 2013. Descriptive metadata for field books: Methods and practices of the Field Book Project. D-Lib Magazine 19 (11/12). https://doi.org/10.1045/november2013-nakasone

Nelson, G., P. Sweeney, L. E. Wallace, R. K. Rabeler, D. Allard, H. Brown, J. R. Carter, M. W. Denslow, E. R. Ellwood, C. C. Germain-Aubrey, E. Gilbert, E. Gillespie, L. R. Goertzen, B. Legler, D. B. Marchant, T. D. Marsico, A. B. Morris, Z. Murrell, M. Nazaire, C. Neefus, S. Oberreiter, D. Paul, B. R. Ruhfel, T. Sasek, J. Shaw, P. S. Soltis, K. Watson, A. Weeks, and A. R. Mast. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. Appl. Plant Sci. 3:1500065. https://doi.org/10.3732/apps.1500065

Palandačić, A., M. Kapun, C. Greve, T. Schell, S. Kirchner, L. Kruckenhauser, N. Szucsich, and N. Bogutskaya. 2024. From historical expedition diaries to whole genome sequencing: A case study of the likely extinct Red Sea torpedo ray. Zool. Scr. 53:32–51. https://doi.org/10.1111/zsc.12632

Peterson, A. T., A. G. Navarro-Siguenza, and R. S. Pereira. 2004. Detecting errors in biodiversity data based on collectors' itineraries. Bull. Br. Ornithol. Club 124:143–151. https://biostor.org/reference/111825

Reznick, D., R. J. Baxter, and J. Endler. 1994. Long-term studies of tropical stream fish communities: The use of field notes and museum collections to reconstruct communities of the past. Am. Zool. 34:452–462. https://doi.org/10.1093/icb/34.3.452

Rinaldo, C., D. Rielinger, J. Deveer, and D. Castronovo. 2023. Connecting libraries, archives, and museums: Collections in support of natural history science. J. Comput. Cult. Herit. 16:1–24. https://doi.org/10.1145/3570905

Sheffield, C., and S. Nakasone. 2011. Together under one roof: Combining collection and item level description through multiple metadata schemas. Proc. Assoc. Inf. Sci. Technol. 48:1–4. https://doi.org/10.1002/meet.2011.14504801312

Simpson, R. C. 2026. Natural history specimens collected and/or identified and deposited. Zenodo. https://doi.org/10.5281/ZENODO.18644268

Stork, L., A. Weber, and K. Wolstencroft. 2019a. The Semantic Field Book Annotator. BISS 3:e37223. https://doi.org/10.3897/biss.3.37223

Stork, L., A. Weber, E. Gassó Miracle, F. Verbeek, A. Plaat, J. Van Den Herik, and K. Wolstencroft. 2019b. Semantic annotation of natural history collections. J. Web Semant. 59:100462. https://doi.org/10.1016/j.websem.2018.06.002

Stork L., A. Weber, J. Van Den Herik, A. Plaat, F. Verbeek and K. Wolstencroft. 2021. Large-scale zero-shot learning in the wild: Classifying zoological illustrations. Ecol. Inform. 62:101222. https://doi.org/10.1016/j.ecoinf.2021.101222

Tingley, M. W., and S. R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends Ecol. Evol. 24:625–633. https://doi.org/10.1016/j.tree.2009.05.009

Webster, M. S. 2017. The extended specimen. Pp. 1–9 *in* The Extended Specimen: Emerging frontiers in collections-based ornithological research. CRC Press, Boca Raton, FL.

Weeks, A. 2020. A new life for Lord Fairfax Community College herbarium. The Vasculum 15:8–12. http://www.herbariumcurators.org/wp-content/uploads/2020/10/Vasculum_15_2.pdf

Weeks, A. 2024. In Memoriam: Lena Artz (1891–1976), a hidden figure of 20th-century southeastern U.S. botany. Castanea 88:297–311. https://doi.org/10.2179/0008-7475.88.2.297

Weeks, A., E. Poindexter, and M. Sheik. 2025. Salvaging the LFCC herbarium: lessons learned about value from a small regional collection. Zenodo. https://doi.org/10.5281/ZENODO.18507185.

Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PLoS ONE 7:e29715. https://doi.org/10.1371/journal.pone.0029715

Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. 2016. The FAIR Guiding Principles for scientific data

management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18

Woller, D. A., and J. G. Hill. 2015. *Melanoplus foxi* Hebard, 1923 (Orthoptera: Acrididae: Melanoplinae): Rediscovered after almost 60 years using historical field notes connected to curated specimens. Trans. Am. Entomol. Soc. 141:545–574. https://doi.org/10.3157/061.141.0301