Echoes from the Heights: Digitizing Specimens from the 1955 Karakoram Expedition

Ashley Bales and Craig Meyer
Philecology Herbarium (herbarium@fwbg.org)
Botanical Research Institute of Texas
at the Fort Worth Botanic Garden

Ashlyn Weaver
Collin County Community College

Introduction

The Himalayan-Hengduan (H-H) mountain range is a hub of unparalleled biodiversity, rivaling that of tropical rainforests, despite the harsh conditions and high elevations. The Karakoram range, in the H-H, is one of the most glaciated parts of the world that has been visited by both outdoor enthusiasts and scientists alike for centuries. However today, the unique species there face critical threats from climate change.

In the summer of 1955, students in the Harvard University Mountaineering Club (HMC) embarked on an expedition to climb the famous peaks in the Karakoram in Pakistan. The accompanying botanist team, Dr. Grady L. Webster and Dr. Eugene Nasir, collected over 900 specimens of vascular plants and bryophytes from altitudes between 7,000-16,000 feet, including several species new to science, and distributed them to herbaria worldwide.

The BRIT Philecology Herbarium, supported by the National Science Foundation-funded All Asia TCN (AATCN) grant[1], is digitizing (imaging, transcribing, and georeferencing) historical collections from H-H and the rest of Asia. Utilizing Webster's archives from the University of California Davis Library[2], alongside modern AI tools, we aimed to fill in the gaps of the HMC 1955 expedition's scientific contributions. This poster highlights the importance of digitization of collections and archives, and shares the challenges and successes of this project.

Archives & AI

Archives

The Herbarium's few collections of Webster's vascular plants from Pakistan have very minimal and sometimes indiscernible or conflicting label data. These specimens alone could only tell us a small part of the expedition's story, leaving staff and volunteers eager to know more about the people behind them. Archives obtained from BRIT Library[3] and UC Davis Library filled in some of the gaps in our knowledge regarding the environmental, social, and historical context of the Karakoram range in 1955.

The BRIT Library archives of Dr. Lloyd H. Shinners, then director of SMU Herbarium, revealed letters to botanists R. R. Stewart and Eugine Nasir at Gordon College in Pakistan in the 1950-1960s. Shinners exchanged college-level botany books with Stewart for the 77 herbarium specimens collected on this 1955 expedition. Webster's archives, shared by the UC Davis Library, included hand-written journal entries, his complete collection number range with updated determinations, newspaper clippings about the the expedition, a botanical report of the explored Pakistan areas, notes on observations of native peoples, and reprint of "The Vegetation and Flora of the Hushe Valley" which included a drawn map of the Ghondokoro and Chogolisa Glaciers in the Karakoram range by expedition surveyor John F. Noxon (Fig 5.).

Machine transcription

Scanned paper images and text are often a huge obstacle for researchers and curators. Though image and text generation gets most of the attention, Large Language Models (LLMs) provide a variety of new tools for scanning images and parsing unstructured text. For this project, the UC Davis's Webster archives provided the only complete list of specimens in a scanned copy of a typewritten list (Fig. 6). OpenAI's recently released GPT-4o engine[4] can handle images as well as text-making a first pass at transcription a matter of seconds rather than hours (Fig. 7). Though admittedly imperfect, this process shifts human attention and time away from rote work and towards QC and editing.

Acknowledgements

We'd like to thank Google Arts & Culture, the National Science Foundation, and FWBG Volunteers for helping with all aspects of digitization for this project and making it possible. We'd also like to thank staff at the University of California Davis Library for sharing Dr. Grady L. Webster's archives and contributing to our specimens' "extended network" of invaluable data.

Fig. 1. (Left) Image of students in the Harvard Mountaineering Club on the side of a mountain in the Karakoram range in Pakistan.
Fig. 2. (Right) Image of a SMU herbarium specimen sheet (BRIT613002)collected on the HMC 1955 expedition.

PLANTS OF KASHMIR
Baltistan

Sophora alopecuroides L.

Abundant in deep sand on high banks above the Indus Rives at Skardu, alt. c. 7500 ft.; flowers yellowish-white.

No. 5766 24 June 1955

Collected by Grady L. Webster & Eugene Nasir

Harvard Mountaineering Club Karakoram Expedition (auspices of Harvard Museum of Comparative Zoology).

Fig. 3. Close-up of BRIT613002 specimen label of *Sophora alopecuroides* L. collected by Grady L. Webster (#5766) and Eugene Nasir on June 24, 1955 in Pakistan.

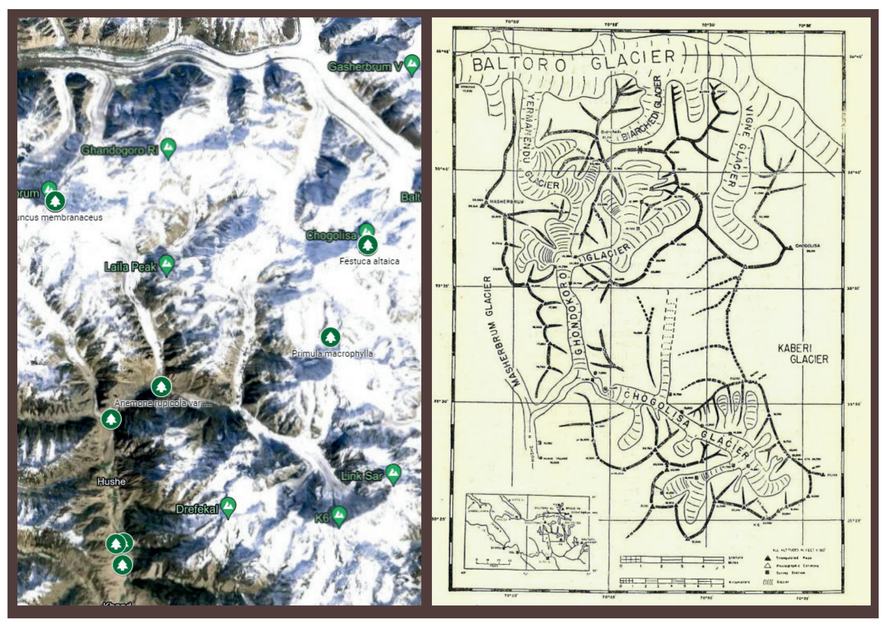

Fig. 4. (Left) Screenshot MyMaps of georeferenced HMC expedition specimens shown on a satellite view of glaciers in the Karakoram range. **Fig. 5.** (Right) John F. Noxon's hand-drawn map of Ghondokoro and Chogolisa Glaciers. Reprinted from "The Vegetation and Flora of the Hushe Valley" by G. L. Davis, 1965, Pakistan Journal of Forestry, Vol. XV, No. 3, p.p. 202. Reprinted with permission from UC Davis Library.

6515 Ferula jaeschkeana Vatke
6 Artemisia salsoloides Willd.
7 Scabiosa speciosa Royle
8 Saussurea falconeri Hook. f.
9* Jurinea macrocephala (DC.) Benth.
20 Heracleum candicans Wall. ex DC.
1 Anemone tetrasepala Royle
2 Lactuca rapunculoides Clarke
3 Erigeron ellisii Hook. f.
4* Delphinium speciosum M. B. var. ranunculifolium (Wall.) Huth
5 Bupleurum longicaule himalayense
6 Bupleurum thomsonii Cl.
7 Silene

Fig. 6. (Above) Example of Webster's archived list of collections from the HMC 1955 expedition.
Fig. 7. (Right) OpenAI GPT-4o engine's transcribed results from archives.

USER

ASSISTANT

Here is the parsed data in CSV format:

```csv
number,scientific name,authority
6515,Ferula jaeschkeana,Vatke
6,Artemisia salsoloides,Willd.
7,Scabiosa speciosa,Royle
8,Saussurea falconeri,Hk. f.
9*,Jurinea macrocephala,(DC.) Banth.
20,Heracleum candicans,Wall. ex DC.
1,Anemone tetrasepala,Royle
2,Lactuca rapunculoides,Clarke
3,Erigeron ellisii,Hk. f.
```

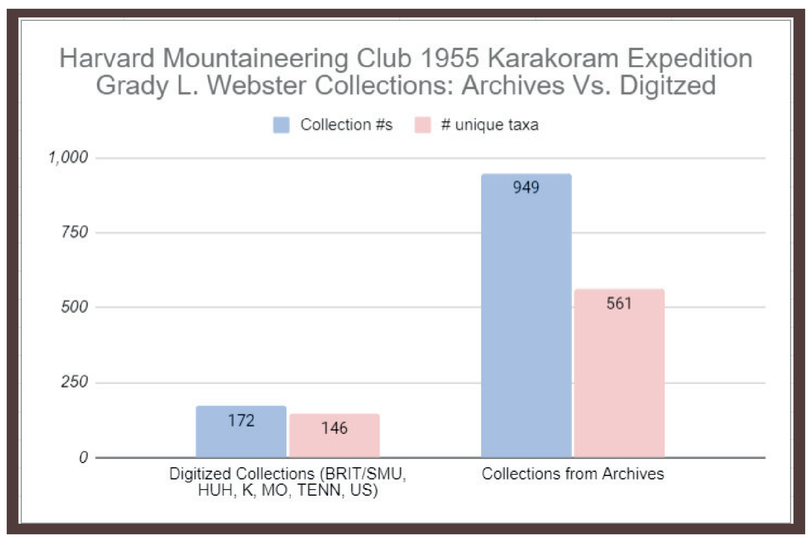

Fig. 8. Bar chart displaying Webster's collection numbers and associated identified unique vascular and bryophyte species that are discoverable through digitized collections compared to those documented in Webster's personal archives.

Digitization Workflow

Imaging

Digitizing the HMC 1955 expedition is part of the Philecology Herbarium's wider effort to digitize specimens under the AATCN grant. Separately, in 2019, BRIT partnered with Google Arts & Culture's initiative to image the holdings of cultural institutions and share them online for global audiences[5]. Google provided the herbarium with a Phase One IQ3 100 megapixel camera that generated images for over 17,000 specimens from across the Asian continent.

Transcription

The Herbarium shares its specimens' data and associated images online through Symbiota's open access TORCH (Texas Oklahoma Regional Consortia of Herbaria) Portal[6]. The Karakoram expedition collections were discovered through the initial efforts of staff and volunteer *skeletal transcription*, capturing only scientific name and country, of all AATCN records. Subsequent and directed *complete transcription* of the Pakistan records captured the remaining data on the specimen labels. This step proved to be particularly difficult in unfamiliar localities; the spelling and even legitimacy of locations require time-consuming research. Keeping and sharing notes on duplicated localities helped to avoid repeating efforts.

Georeferencing

The final step, *georeferencing*, uses the specimen label's provided locality data to estimate latitude, longitude, and uncertainty radius using the GEOLocate web-based collaborative client tool[7]. Georeferencing can be intimidating, and the most challenging part is overcoming the fear of being incorrect about the placement of the georeferenced point. Interpreters want to be right, and that's not always possible when locality data can often be inaccurate and require additional research. Aiming for "defensible interpretation" rather than "absolute certainty" is the best way to ease individual anxiety. However, this fear is challenged by providing each locality with a strong amount of justification for the interpretation and any sources used for future geolocating. For the AATCN, staff and volunteers georeferenced 333 specimen records from Pakistan, including those from the HMC 1955 expedition (Fig. 4).

Conclusions

Herbarium specimens provide a snapshot in time of native flora and can be important data tools for species in threatened environments like the H-H mountain range. Digitization of these specimens and sharing images and associated data online can be invaluable to researchers. For this project focusing on the Harvard Mountaineering Club expedition of the Karakoram range, archives provided by the University of California Davis Library (with help from OpenAI) proved to be an important piece to the complete botanical surveying and collection efforts in the summer of 1955 by Dr. Grady L. Webster and Dr. Eugene Nasir. Of the 949 specimens (Webster #s 5650 to 6599) of bryophytes and vascular plants collected, only 218 (including duplicates) have been digitized and made discoverable through online portals such as TORCH and GBIF by these herbaria of the deposited collections: HUH, K, MO, TENN, and US (Fig. 8). This leaves more than 730 collections undigitized and potentially unprocessed at the herbaria of deposit, and thus leaving the biodiversity data of the region hidden. The processing and digitizing collections from biodiversity hotspots around the world and associated archives by botanists should be seriously considered as a priority as we face critical threats of climate change in our future.

References Cited

1 All Asia TCN: NSF Award No. 2101846
2 University of California Davis Library, Archives and Special Collections: library.ucdavis.edu/archives-and-special-collections/
3 Botanical Research Institute of Texas Library and Special Collections: fwbg.org/research/library-special-collections/
4 OpenAI. (2024). ChatGPT (May 13 version) [Large language model]. chat.openai.com/chat
5 Google Arts & Culture: artsandculture.google.com
6 TORCH Data Portal: portal.torcherbaria.org/
7 Collaborative Georeferencing Portal for GEOLocate: coge.geo-locate.org/