BarnebyLives an R package to create herbarium specimen labels

Reed Clark Benkendorf[1]*, Nyree Zerega[1,2], Pat Herendeen[1,2], Jeremie B. Fant[1,2]

[1]Chicago Botanic Garden, 1000 Lake Cook Road, Glencoe, Illinois 60022, USA

[2]Plant Biology and Conservation, Northwestern University, Evanston, Illinois 60208, USA

Abstract

Premise: Depositing quality specimens to herbaria is a time intensive task. Many institutions have reduced the amount of funding for herbaria, and universities have reduced the amount of education dedicated to curatorial tasks and specimen deposition. Despite this, the continual generation of herbaria specimens are essential for research in ecology and evolution. In order to faciliate the continued growth of herbaria BarnebyLives was developed as tool to supplement collection notes, perform geographic and, taxonomic informatic processes, enact spell checks and produce labels.

Methods and Results: BarnebyLives uses geospatial data from the U.S. Census Bureau to provide political jurisdiction information, and data from other sources, including the United States Geological Survey, to supplement collection notes by providing information on abiotic site conditions. It uses inhouse spell checks to verify the spelling of a collection at all taxonomic ranks, the IPNI standard author database to check standard author abbreviations, and the Royal Botanic Garden Kews 'Plants of the World Online' to check for nomenclatural innovations. Optionally the package writes driving directions to sites using Google Maps. Finally the package outputs data in a tabular format for review by the user to accept or confirm changes,

Conclusions: BarnebyLives provides accurate political and physical information, reduces typos, provides users the most current taxonomic opinions, generates driving directions to sites, and produces aesthetically appealing labels and shipping manifests in a matter of minutes.

Nearly 400 million specimens are housed in herbaria around the world (Thiers 2019). These specimens were collected with the goal of describing the plant kingdoms taxonomic diversity, and documenting the worlds floristic diversity (Greve et al. 2016). The rate of accessioning new collections to herbaria diminished in the 20th centry as research goals in the biological sciences shifted away from describing, documenting, and understanding earths biodiversity (Prather 2004, Pyke & Erlich 2010). Which, among other factors, lead to a

---

*Correspondence: rbenkendorf@chicagobotanic.org

decline in the amount of funding allocated to collections based research, and the number of staff maintaining and accessioning new collections (Funk 2014). Fortunately, renewed interest in collections have brought herbaria back to the forefront of plant sciences (Ronsted et al. 2020).

Recent innovations in computing, specimen digitization, data sharing, DNA sequencing, and statistics have brought about a renaissance in herbarium based studies (Greve et al. 2016, James et al. 2019, Brewer et al. 2019, Ronsted et al. 2020). Current uses of specimen based data extend far beyond their traditional roles in systematics and floristics, and studies utilizing collections are regularly carried out to better understand . . . . of plants (). However, we anticipate that collections will gain their most widespread utilization as natural history is being revitalized in ecology, via novel approaches, such as remote sensing, meta-barcoding, and . . . (Fevo et al. 2021).

However, we now stand at a time where we recognize the need for more specimens, but are in a difficult position where the skills of collecting and processing specimens have declined amongst young persons (Daru et al 2017). The submittal of specimens to herbaria is a, well documented albeit time consuming process, especially for younger collectors with limited experience in the process. While many young collectors, who are capable of using dichotomous keys to reliably identify their collections, exist we have observed that they face difficulties navigating several aspects of data collection. This scenario results in not only the delay in the deposition of many specimens, but undoubtedly the deposition of many collections at all. Problems which young collectors face generally include the lack of dedicated time awarded to them at a seasons end to process specimens, and the lack of formal education on cartography, natural history, and taxonomy at most universities.

The successful generation of an herbarium specimen includes many steps which are easy to take for granted. For example, while the acquisition of political information for a collection site appears simple, it is only so if the collector has the adequate resources at their disposal. Given the association of boundaries with topographically complex areas (e.g. watersheds) it often requires topographic maps, which are no longer widespread - resulting in many having difficulties interpreting them, or transcription of coordinates into a Geographic Information System (e.g. ArcMap, which is relatively expensive at 100$ year), or more likely Google Maps by individual site. This lack of topographic maps compounds the issues of young collectors being unable to come up with appropriate site names.

More evidently difficult tasks involve taxonomy and the rapid rate at which taxonomic names have changed since the publication of many Floras.

METHODS AND RESULTS

Here we showcase the reduction in errors attributtal to BarnebyLives

CONCLUSIONS

AUTHOR CONTRIBUTIONS

The project was conceptualized by R.C.B. The program was written by R.C.B Data collection and analysis were performed by R.C.B. R.C.B. wrote the manuscript with input from all other authors. All authors approved the final version of the manuscript.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The BarnebyLives R package is open source, the development version is available on GitHub (https://github.com/sages teppe/BarnebyLives), and the stable version is available on CRAN. The package includes three real use-case vignettes (tutorials) on usage. One vignette "setting_up_files" explores setting up a instance for a certain geographic area. Another vignette "running_pipeline" showcases the usage of the package for processing data entered on a spreadsheet. A final vignette "creating_labels" shows the usage of an R, and Bash script launched from RStudio to produce print-ready labels.
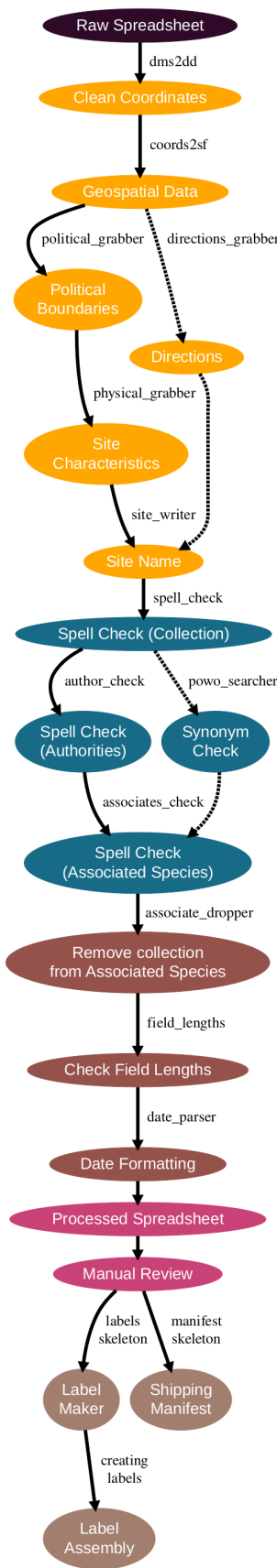

Figure 1: Specified benchmarks at all ESDs

ORCID

Jeremie Fant https://orcid.org/0000-0001-9276-1111

REFERENCES