BarnebyLives an R package to create herbarium specimen labels

Reed Clark Benkendorf[1]*, 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 (n.d.)). 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 et al. (2004), Pyke and Ehrlich (2010)). Which, among other

---

*Correspondence: rbenkendorf@chicagobotanic.org

factors, lead to a 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 (Rønsted 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. (2018), Brewer et al. (2019), Rønsted 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 . . . (Tosa 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.
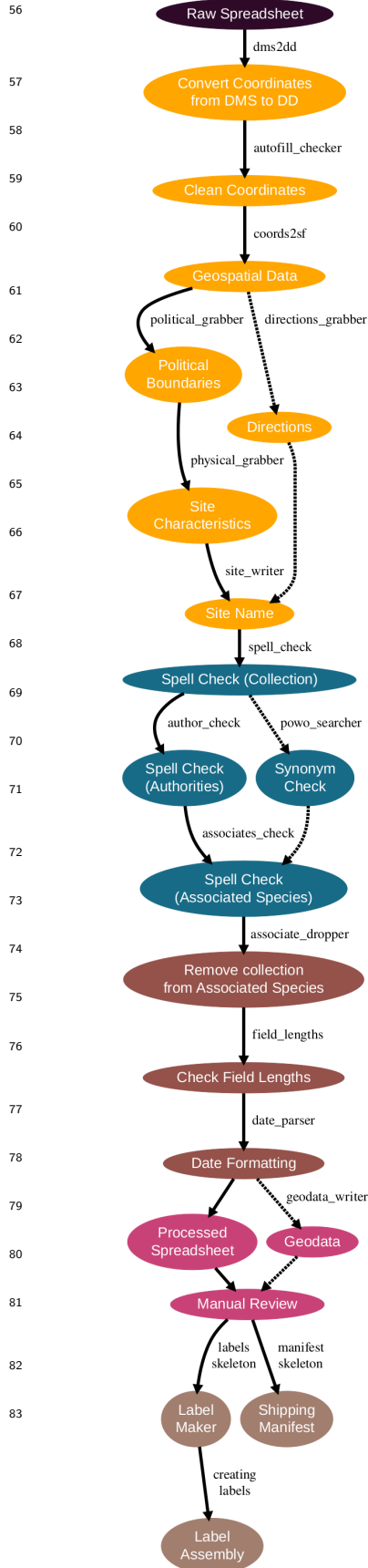
Figure 1: Specified benchmarks at all ESDs which included all three metrics and derived imputed value

METHODS AND RESULTS

Sit massa risus egestas vivamus condimentum fermentum sapien velit est, semper nec risus. Porta taciti; morbi ante fames molestie cubilia senectus nibh platea sagittis. Sociosqu litora aliquam ligula convallis ridiculus dapibus, vel – in etiam bibendum mollis fringilla.

Sit mus class, penatibus ante interdum primis nisl tempus! Inceptos risus faucibus duis cursus a orci, lacinia scelerisque taciti commodo congue hendrerit ad. Penatibus dignissim primis, tempus in tortor, bibendum maecenas. Himenaeos morbi per – turpis enim, eu porttitor laoreet mattis ornare elementum. Feugiat vehicula etiam porttitor, sem eu quam mauris nunc, sodales, malesuada vivamus risus porta.

Dolor sociosqu proin ad sem! Dis diam lectus aptent: nascetur convallis gravida quam ultricies semper. Ridiculus nascetur nullam curabitur cras cras semper ligula montes justo suspendisse. Suspendisse aenean nascetur quam quis: id tristique – montes dis eu parturient ornare congue massa posuere viverra facilisis.

Amet proin aenean, maecenas gravida odio, congue imperdiet. Ullamcorper orci ante nascetur aptent rutrum a risus varius nullam facilisi fringilla. Commodo neque mus fringilla posuere praesent vel eros gravida eget suscipit, ad ut facilisi. Tristique lacus tempor commodo fermentum etiam; bibendum vehicula?

Lorem himenaeos sem, nostra torquent sed, cum ante? Ligula nisi facilisis phasellus potenti tellus, ad sem condimentum. Tempus nisl sapien montes aptent – cursus ornare gravida. Iaculis porttitor imperdiet aptent volutpat est netus in: dictumst porttitor libero. Commodo nisl habitant elementum!

Here we showcase the reduction in errors which are associated with the implementation of BarnebyLives.

3

CONCLUSIONS

Ipsum posuere egestas himenaeos ligula nec dignissim. Nullam ac tempor iaculis lacus montes sollicitudin, platea consequat massa. Ad eros erat arcu aliquam eleifend mattis, sociosqu mauris fames taciti! Malesuada in inceptos mollis iaculis scelerisque semper a, felis cras cubilia pulvinar!

Adipiscing sem vulputate, quisque habitasse egestas turpis? Cras aliquet himenaeos; facilisis inceptos eu imperdiet vehicula magnis sollicitudin mus dis? Taciti fusce eros commodo netus laoreet aliquam in vivamus! Habitant rhoncus senectus potenti mus cubilia litora bibendum praesent interdum! Interdum purus aliquet – sem nibh erat nisl. Phasellus nullam lacus egestas vivamus blandit auctor viverra tortor metus viverra sed varius vulputate.

Consectetur quisque sagittis sollicitudin mollis leo tristique, molestie, aptent velit. Augue cubilia mollis vehicula ornare eleifend nascetur, aenean nam nam. Sed mattis ultricies laoreet curabitur sed pretium. Cras velit ad dis magnis, bibendum ligula, cras faucibus semper magnis platea: pulvinar: eleifend, est penatibus.

Data Sources for Package					
Variable	Usage	Source	Name	Data Model	Size (GiB)
County	Political	US Census Bureau	Counties	Vector	0.073
State			States		0.0*
Ownership		US Geological Survey	Protected Areas Database		0.435
TRS			Public Land Survey System		0.816
Place Names	Site Name		Geographic Names Information System		0.081
Mountains	Site Name	EarthEnv	GMBA Mountain Inventory v2		0.004
Elevation	Site Characteristics	Open Topography	Geomorpho90m - Elevation	Raster	4.2
Slope			Geomorpho90 - Slope		4.6
Aspect			Geomorpho90m - Aspect		4.1
Geomorphons			Geomorpho90m - Geomorphons		0.455
Surficial Geology		US Geological Survey	State Geologic Map Compilation	Vector	0.708
Taxonomic Spellings	Spell Checks	World Flora Online	World Flora Online	Text	0.002
Author Abbreviations		IPNI	International Plant Names Index		0.001
*Counties and States are merged into the same dataset while setting up the package. The value for "County" includes State.					

Figure 2: Sources of Data required for operations

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.

4

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

The BarnebyLives R package is open source, the development version is available on GitHub (https://github.com/sagesteppe/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.

ORCID

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

REFERENCES

Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in plant science* 10: 1102.

Funk, V. A. 2014. The erosion of collections-based science: Alarming trend or coincidence. *The Plant Press* 17: 1–13.

Greve, M., A. M. Lykke, C. W. Fagg, R. E. Gereau, G. P. Lewis, R. Marchant, A. R. Marshall, et al. 2016. Realising the potential of herbarium records for conservation biology. *South African Journal of Botany* 105: 317–323.

James, S. A., P. S. Soltis, L. Belbin, A. D. Chapman, G. Nelson, D. L. Paul, and M. Collins. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. *Applications in plant sciences* 6: e1024.

Prather, L. A., O. Alvarez-Fuentes, M. H. Mayfield, and C. J. Ferguson. 2004. The decline of plant collecting in the united states: A threat to the infrastructure of biodiversity studies. *Systematic Botany* 29: 15–28.

Pyke, G. H., and P. R. Ehrlich. 2010. Biological collections and ecological/environmental research: A review, some observations and a look to the future. *Biological reviews* 85: 247–266.

Rønsted, N., O. M. Grace, and M. A. Carine. 2020. Integrative and translational uses of herbarium collections across time, space, and species. *Frontiers in Plant Science* 11: 1319.

Thiers, B. M. The world's herbaria 2021: A summary report based on data from index herbarium.

Tosa, M. I., E. H. Dziedzic, C. L. Appel, J. Urbina, A. Massey, J. Ruprecht, C. E. Eriksson, et al. 2021. The rapid rise of next-generation natural history. *Frontiers in Ecology and Evolution* 9: 698131.