Tracking Data from Download to Publication - and Back?

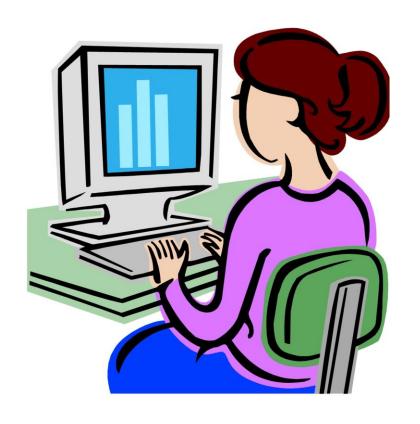
Pamela S. Soltis
Florida Museum of Natural History
UF Biodiversity Institute
iDigbio
University of Florida





















Tracking use of collections data...

• From download to use to publication - and back



What should we track?

What are the challenges, and how can they be overcome?

Tracking use of collections data...

- From download to use to publication and back
 - Not necessarily a simple path



What should we track?

What are the challenges, and how can they be overcome?

Tracking use of collections data...

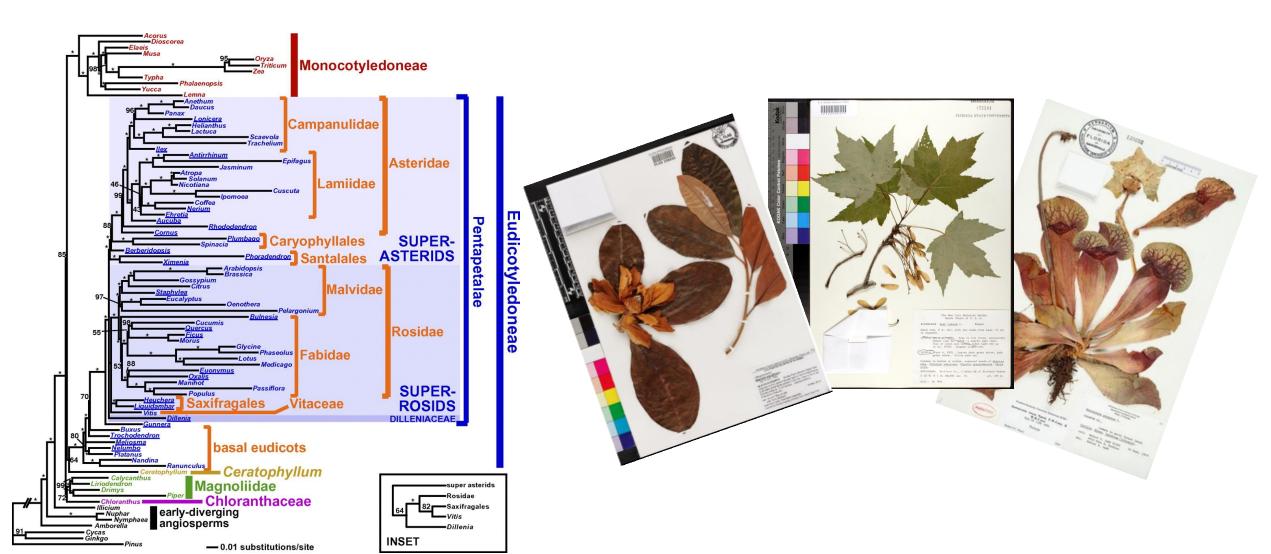
- From download to use to publication and back
 - Not necessarily a simple path
 - No consensus on what (if anything) should be done
- Why track data?
- What should we track?

What are the challenges, and how can they be overcome?

Why track data?

- Transparency
 - Open, reproducible science
 - Opportunities for error correction and annotation
- Allows for eventual integration with other data
 - Augmentation of records GenBank, traits, etc.
 - Community resource
- Attribution
 - Collectors, institutions, curators, data managers, et al.

Vouchers...



Citing museum specimens...

Anthocyanidin Variation in Clarkia

PAMELA S. SOLTIS

Department of Botany, Washington State University, Pullman, WA 99164, U.S.A.

Key Word Index—Clarkia; Onagraceae; anthocyanidins; intersectional relationships; chemosystematics.

Abstract—Previous reports of anthocyanin variation in *Clarkia* indicated intersectional differentiation, suggesting an additional marker for evaluating infrageneric relationships. Therefore, an investigation of anthocyanidin content of 19 species representing seven of the 10 sections was conducted. The anthocyanidins malvidin, cyanidin and delphinidin were detected. Interspecific variation in anthocyanidin profile was observed: some taxa possessed only malvidin, whereas others exhibited all three anthocyanidins. However, this variation does not coincide with proposed sectional boundaries based on evidence from morphology and enzyme electrophoresis. This suggests that anthocyanidins are homoplasious characters in *Clarkia*.







Experimental

Nineteen species of *Clarkia* were examined for anthocyanidin variation. Floral anthocyanin samples were collected from living plant material of 18 populations representing 13 species (see below). Samples were also obtained for an additional eight species from herbarium specimens (WS). Extraction and hydrolysis of anthocyanins, thin-layer chromatography and identification of anthocyanidins were conducted following the procedures of Dorn and Bloom [7] and Soltis and Soltis [9].

Collection locations. Section Godetia: Clarkia purpurea, California. Tulare Co.: 0.5 mi E of junction with road to Glennville on White River Rd.; Madera Co.: Rd. 211, 4.35 mi SSW of junction with Rd. 210. C. williamsonii, California. Madera Co.: 20.9 mi S of Rock Creek Campground on Mammoth Pool Rd.; Hwy. 41, 2.1 mi N of junction with Hwy. 49; Tulare Co.: Rd. 245, 1.0 mi N of junction with Rd. 465. C. davyi, California. Mendocino Co.: Russian Gulch State Park; Hwy. 1, S of Ft. Ross Historic Site. Section Peripetasma: C. cylindrica, California. Kern Co.: Rd. 226; Caliente-Bodfish Rd. C. dudleyana, California. Kern Co.: mile-marker 46.5 on Hwy. 155; Tulare Co.: Rd. 245, 1.0 mi N of junction with Rd. 465. C. biloba, California. Mariposa Co.: Hwy. 49, 12.3 mi W of junction with Hwy. 140. Section Phaeostoma: C. xantiana, California. Kern Co.: Rd. 483; Breckenridge Mtn. Rd. C.

Herbarium specimens (WS). Section Clarkia: Clarkia pulchella. Section Eucharidium: C. breweri, C. concinna. Section Myxocarpa: C. rhomboidea. Section Peripetasma: C. lewisii, C. modesta. Section Rhodanthos: C. arcuata, C. lassenensis.

Modeling the distribution of species

- Location information and environmental data
- Maxent to model the range of each species
- For Florida plants:
 - 1,490 plant species (of 4100 species)
 - >511,000 georeferenced points
 - Environmental features: temperature, precipitation, soil, etc.



Julie Allen



Charlotte Germain-Aubrey



What should we track?

- Downloaded data
- Filtered data
- Cleaned data
- Analyzed data
- Etc.
- e.g. via
 - GUIDs (multiple)
 - doi, per GBIF
 - NOT in supplements, pdfs, etc.











Citation guidelines

These guidelines provide the most common examples of citation by GBIF users.



Chicory (Cichorium intybus) by Donald Hobern. Photo licensed under CC BY 4.0.

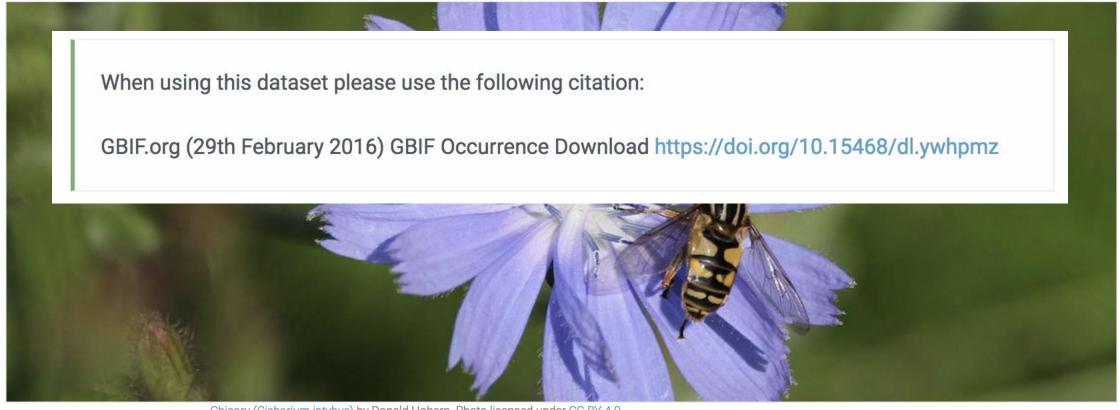






Citation guidelines

These guidelines provide the most common examples of citation by GBIF users.



Chicory (Cichorium intybus) by Donald Hobern. Photo licensed under CC BY 4.0.

Dicerandra



Andre Naranjo





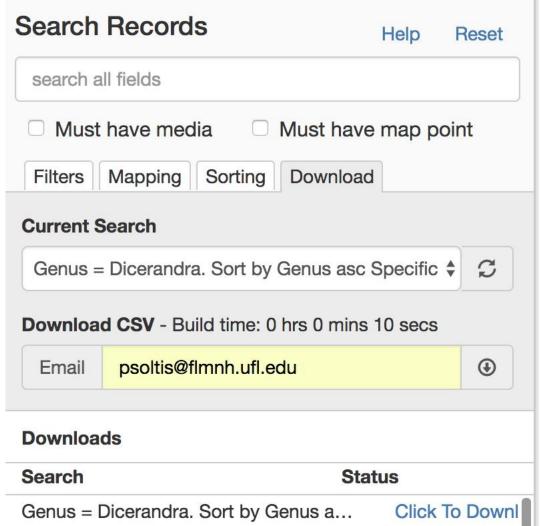


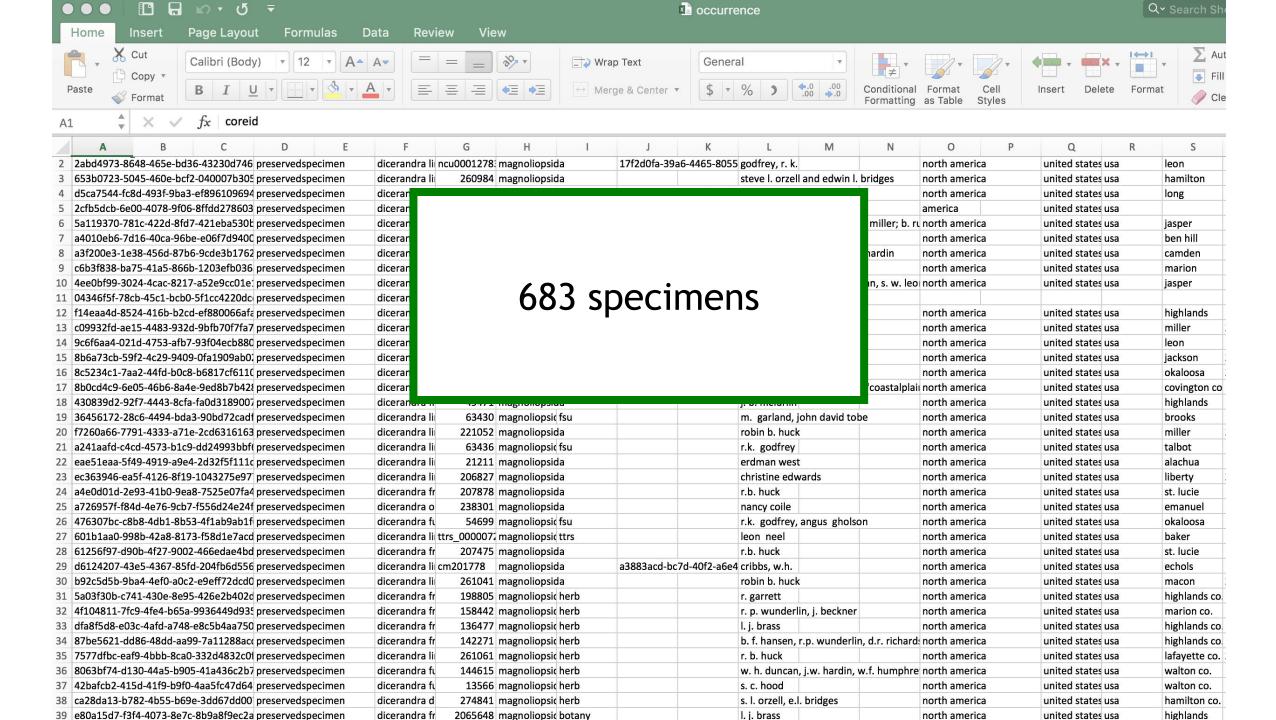














Scientific Name

Dicerandra densiflora

Family

Lamiaceae

Date Collected

1975-10-01

Must have map point...

About iDigBio Research Technical Information Education

Basis of Record

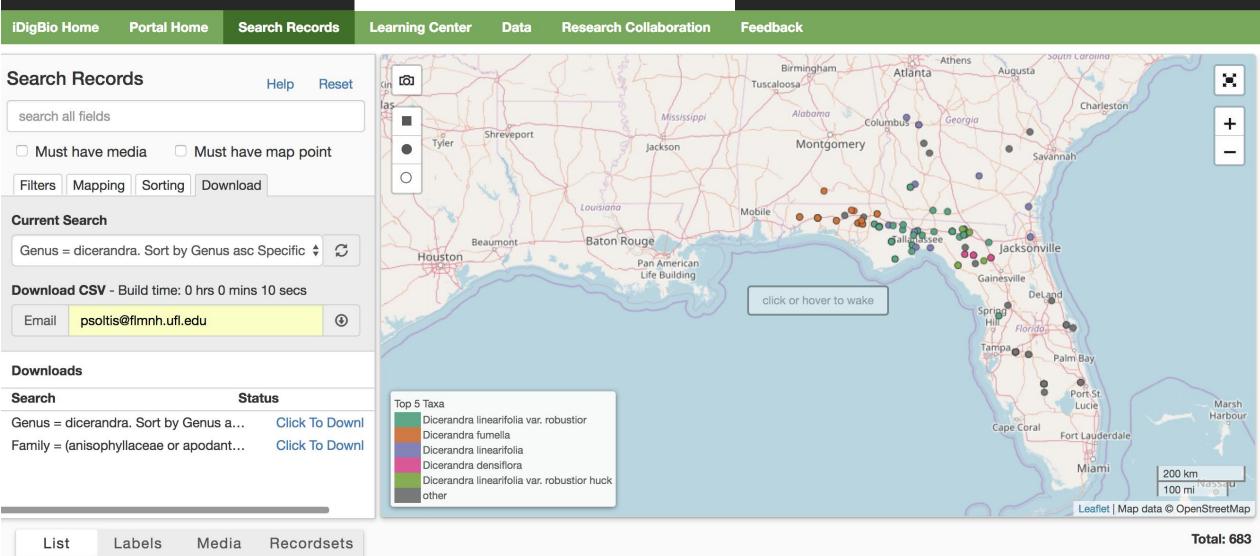
PreservedSpecimen

Columns

view

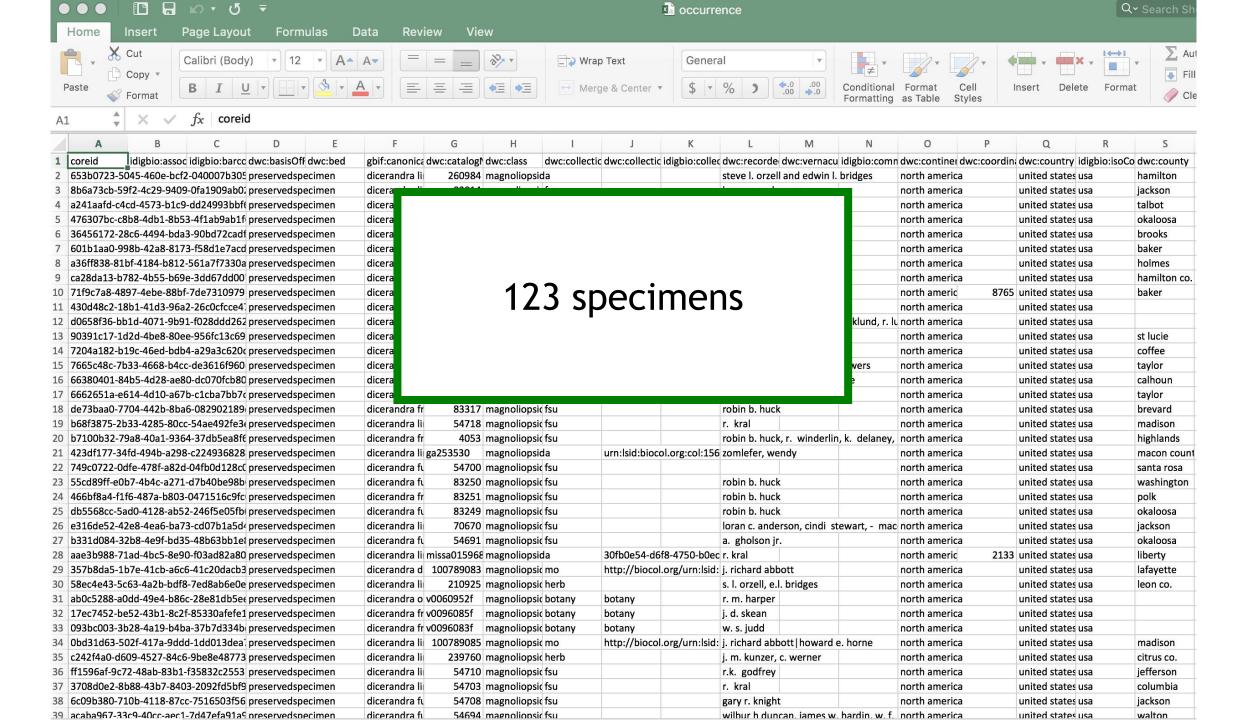
Institution Code

FLAS



Country

United States





Lamiaceae

Dicerandra linearifolia

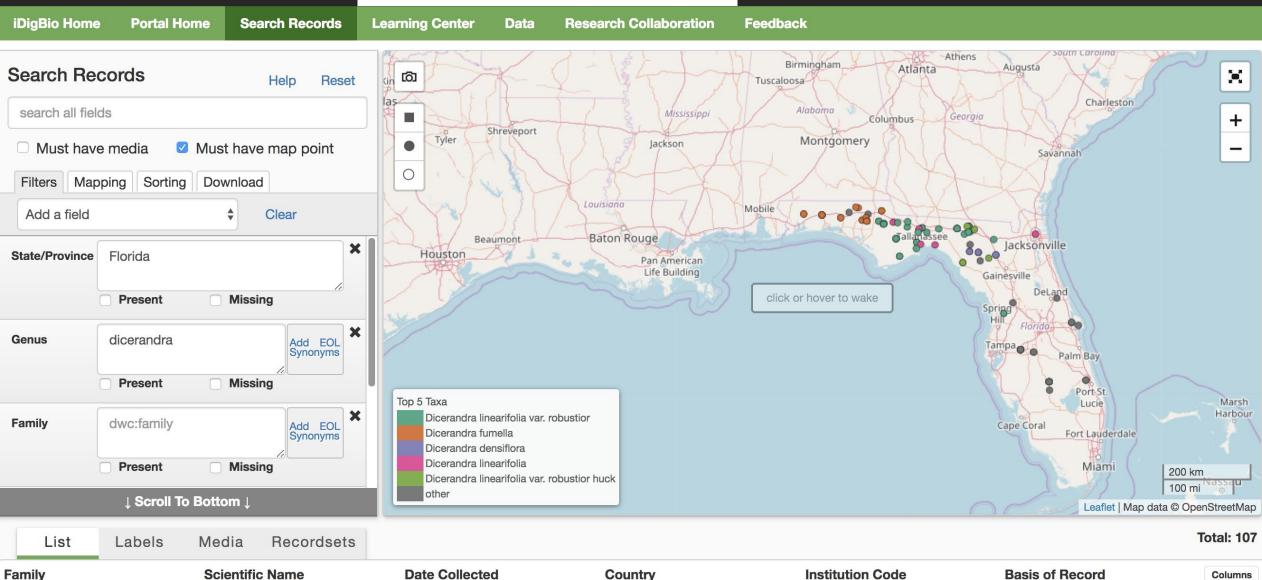
no data

Must have map point... & be in Florida

About iDigBio Research Technical Information Education

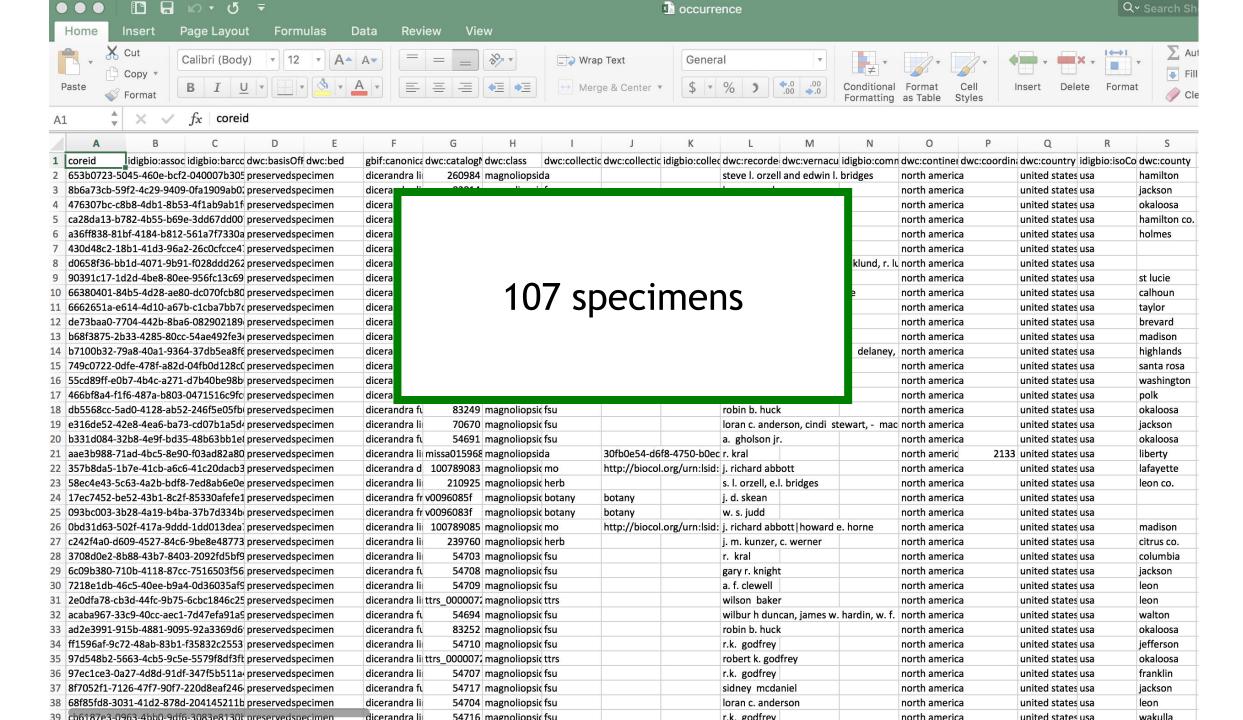
PreservedSpecimen

view



United States

TTRS



Dicerandra



Andre Naranjo













If data sets were all published...

- Clear record of data used & how
- Annotated specimen records
- Reproducible analysis
- Data could be integrated with other data, augmentation of records
 - Genetics
 - Functional traits
 - Env. data, e.g. NEON
- Attribution
 - Institutions can search for their data
 - All institutions, collections that contributed value get credit

Why don't we track and publish data sets?

- No consensus
 - What to track
 - How to track
 - Where to publish
- Technical challenges: how to cite, and where to publish?
 - GUIDs
 - GBIF doi model
 - Dryad, others
- Social
 - Submit to publishers that will accommodate data sets
 - Need to convince 'journals' that data should be published
 - 'we' are the journals: reviewers, editors, editorial boards, etc.

Acknowledgments

- iDigBio
- NSF
- Julie Allen, Charlotte Germain-Aubrey, Narayani Barve, Andre Naranjo
- Photo credits: L. Elliott, M. Harms, E. Warne, D. Bender, Wikimedia



