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

- Why track data?

- 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

• Why track data?

• 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 Dom and Bloom [7] and Soltis and Soltis [9].

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.
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.
Citation guidelines

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

When using this dataset please use the following citation:

GBIF.org (29th February 2016) GBIF Occurrence Download https://doi.org/10.15468/dl.ywhpmz

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

Andre Naranjo

---

**Search Records**

- **Genus** = Dicerandra. Sort by Genus asc

**Download CSV** - Build time: 0 hrs 0 mins 10 secs

**Email** psoltis@flmnh.ufl.edu

**Downloads**

<table>
<thead>
<tr>
<th>Search</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genus = Dicerandra. Sort by Genus asc</td>
<td>Click To Download</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>683</td>
</tr>
</tbody>
</table>

The document contains a table listing various specimens with their details such as collection numbers, descriptions, and localities.
Must have map point...
| A   | B      | C                | D                | E                | F                | G                | H                | I                | J                | K                | L                | M                | N                | O                | P                | Q                | R                |
|-----|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
|     | coreid | ddb accession   | dwc class        | dwc record       | dwc vernacular  | ddb commune      | ddb country      | ddb iso            | ddb county         |                  |                  |                  |                  |                  |                  |                  |                  |
| 123 | specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens | 123 specimens |

123 specimens
Must have map point... & be in Florida
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>specimens</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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