The Digitization Experience at Florida State University’s Robert K. Godfrey Herbarium

Austin Mast
Director, Robert K. Godfrey Herbarium
Associate Professor, Department of Biological Science
Florida State University
Outline

• Intro to the herbarium
• Our history of digitization
• The front-end
• Notes from Nature Project
• General observations
• General recommendations
Robert K. Godfrey (1911–2000)
Godfrey was a prolific collector.
Portion of 76,577 digitized specimens (of ca. 210,00 total) collected by herbarium directors.

- Godfrey: 33%
- Anderson: 21%
- Others: 46%
- Kurtz (<1%) and Mast (<1%)
The herbarium is arguably the most extensive plant collection documenting this biotic hotspot: a region with ca. 2864 spp. of native vascular plants. 125 plant taxa are endemic to it.
The herbarium is arguably the most extensive plant collection documenting this biotic hotspot: a region with ca. 2864 spp. of native vascular plants. 125 plant taxa are endemic to it.
The bulk of the Microalgae Collection originates from samples of phytoplankton in Florida’s coastal and inland waters. In addition to plankton samples, the collection also includes benthic diatoms that are associated with submersed aquatic plants (epiphytic), sediments (epipellic), and artificial substrates, such as glass slides (periphytic). These samples were collected by Professor Robert J. Livingston and Dr. A. K. S. K. Prasad of FSU’s Department of Biological Science over a period of 25 years. Florida collection localities include the drainage basins of the Perdido, Choctawhatchee, Chipola, and Suwannee Rivers, the Econfina and Fenholloway estuaries, and Florida Bay, among others. The collection also includes specimens from elsewhere along the Gulf and Atlantic coasts of the southeastern US, the Caribbean Sea, the Indian Ocean, the Equatorial Pacific (including the Galápagos Islands), and the South China Sea. The collection includes over 8000 permanent diatom slide preparations mounted in Naphrax or Zrax, over 7000 scanning and transmission electron microscope images, and over 40,000 light microscopic images. Extensive research publications and manuals of diatoms (ca. 2000) from the personal libraries of Dr. Greta A. Fryxell (formerly of Texas A&M University), William Miller III (formerly of FSU EM lab), and Dr. Prasad provide added value to these important collections.
Outline

- Intro to the herbarium
- **Our history of digitization**
- The front-end
- Notes from Nature Project
- General observations
- General recommendations
We began work on the database and front-end in Fall 2003.
An important resource for us at the time has HISPID 3 due to its fairly exhaustive representation of specimen qualities. Darwin Core (and DC Extensions) would be a good place to start today.
We are on our second digital camera, but we are still using the same lights, copy stand, and (in some cases) computers.
Number of specimens databased in each of 10 years. Total = 76,577; Range = 3,279 – 14,767; Mean = 7,658.
Digitization has involved 12 graduate students, 59 undergraduate students, and others.
Types of Undergraduate Participation

- Volunteers (*sparingly*)
- Directed Independent Study
- Federal Work Study
- FSU’s Women in Math, Science, and Engineering Program
- FSU-TEACH Program
Types of Undergraduate Participation

Volunteers are typically trained to mount new specimens.
Graduate student participation is by means of a permanent half-time curator line and research assistantships. Funding for just some of the projects shown here included RA support (iDigBio does not).
Graduate student participation is by means of a permanent half-time curator line and research assistantships. Funding for just some of the projects shown here included RA support (iDigBio does not).
The greatest number of specimens digitized occurred with the Deep South Project, which involved imaging first then databasing (hence the lag), and the DEP/FWCC contract. The iDigBio grant does not pay for digitization.
The Tall Timbers grant resulted in a wave of new functionality for the front-end.
Outline

• Intro to the herbarium
• Our history of digitization
• The front-end
• Notes from Nature Project
• General observations
• General recommendations
In August of 2003 the Robert K. Godfrey Herbarium set up a digital imaging system and SQL database. Currently 76,228 of our 200,000+ specimens have been entered into the herbarium's database. In late 2010, Tall Timbers Research Station's Herbarium began serving specimen data and images on this site as well. Currently, data and images for 10,344 of Tall Timbers Research Station's 10,000+ herbarium specimens are being served.
Florida Torrey is restricted to 3 counties in the Florida Panhandle and an adjacent county in Georgia.

Specimen Details

Species: Venicia fordii
Annotated Since Imaged?: True
Collection Date: June 5, 2001
Bar Code ID: 000000945
Collectors: Loran C. Anderson
Collector's Identifier: 19779
Flower Buds Present?: False
Flowers Present?: False
Fruit Present?: True
Country: United States
State or Province: Florida
County or Parish: Leon
Fips Code: 12073
Nearest Named Place: Tallahassee
Verbatim Directions to Locality: bordering the railroad tracks just E of Capital Circle NW in NW sector of Tallahassee.
Latitude - degrees: 0
Latitude - minutes: 0
Latitude - seconds: 0
Latitude - degree with decimals: 0
Longitude - degree: 0
Longitude - minutes: 0
Longitude - seconds: 0
Longitude - degree with decimals: 0
Habitat: woody area bordering the railroad tracks.
Plant Morphology Observations: small tree; all fruits on this tree are 5-carpellate, but a few of them have only 3 or 4 seeds by abortion.
Non digital information?: N

Identification Records

Identifier's Name: Loran C. Anderson
Formal Identification Date: 2001-06-05
Specimen Database Search
Please cite!

Search Criteria

Sort By: 

Show:  
Hide:  

To modify your search choose the "Show" radio button above. If you would like to sort your results differently, change the field in "Sort by" above.

Search Results

Results 1-19 of 19  Page 1 of 1
### Specimen Database Search

**Search Criteria**

- **Institution:** Any Institution
- **Family:** Spacebar yields full listing.
- **Genus:**
- **Species:** Vernica fordi: The scientific name (e.g., Pinus palustris).
- **Common Name:**
- **Collection Date:** (YYYY-MM-DD or MM-DD)
- **Collector Name:**
- **Collector's Identifier:**
- **Barcode:**
- **Country:**
- **State:**
- **County:**
- **Nearest Named Place:**
- **Flowers present:**
- **Fruit present:**
- **Habitat:**

**More Options:** Show: Hide:

**Sort By:** Species

**Output Type:** Map
Specimen Database Search

Search Criteria

Sort By: Species

To modify your search choose the “Show” radio button above. If you would like to sort your results differently, change the field in “Sort by” above.

Search Results

Results 1-19 of 19 Page 1 of 1

Bar Code: 000000556
Family: Euphorbiaceae
Species: Vernicia fordii
Country: United States
State: Florida
County: Franklin
Collector: Anderson, Loran C.
Collector's Identifier: 6944
Collection Date: 1984-04-06
The user can also choose “Report” as output type and download a tab-delimited file to open in Excel.
Outline

• Intro to the herbarium
• Our history of digitization
• The front-end
• Notes from Nature Project
• General observations
• General recommendations
Public participants could help the community digitize its HUGE backlog of specimens.
In this tool, the user first draws a box around the label.

http://www.notesfromnature.org/
The label is then shown blown-up, and the user enters information into 9 fields from it.
These 9 fields were chosen to maximize the specimen’s discoverability and value for a large fraction of specimen-based research.
FSU contributed the first 3531 specimen images for the Notes from Nature herbarium interface. The average number of transcriptions for those specimens was 6.3.

http://www.notesfromnature.org/
Inter-replicate agreement was highest in those starred (*) fields.
Notes from Nature and iDigBio’s CITSCribe Hackathon next week will develop new functionality to enable public participation in the transcription of specimen labels. Remote participants are welcome.
Outline

- Intro to the herbarium
- Our history of digitization
- The front-end
- Notes from Nature Project
- General observations
- General recommendations
Many of FSU’s specimens were collected by FSU researchers and don’t appear to be widely duplicated in collections with aggressive digitization programs.
Small herbaria, such as those involved in the Deep South Imaging Project, do not appear to share a lot of duplicated specimens.
Small herbaria, such as those involved in the Deep South Imaging Project, do not appear to share a lot of duplicated specimens.
And those small herbaria can provide complementary information to build a more complete picture of diversity through time.
Our ability to tackle taxonomically defined priorities began to diminish as the need to digitize new accessions of taxa in folders already stamped “databased” increased. Here is an example of an early undergraduate digitization project.

**Project: Citrus of Florida**

**Project Leader:** Danny F. LaBarbera

**Project Description:**
Florida specimens of the family Rutaceae.

**Project Species (Family, Genus, Species):**

<table>
<thead>
<tr>
<th>Family</th>
<th>Species</th>
<th>Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rutaceae</td>
<td>Amyris balsamifera</td>
<td>BALSAM TORCHWOOD</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Amyris elemifera</td>
<td>SEA TORCHWOOD</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Amyris madrensis</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Amyris texana</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Choisya dumosa</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Citrus aurantium</td>
<td>GRAPEFRUIT, SOUR ORANGE, SWEET ORANGE</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Citrus depressa</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Citrus limon</td>
<td>LEMON</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Evodia nishimurae</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Geijera linearloba</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Orixia japonica</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Phellodendron amurense</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Poncirus trifoliata</td>
<td>HARDY ORANGE</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Ptelea augustifolia</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Ptelea crenulata</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Ptelea trifoliata</td>
<td>COMMON HOPTREE, WAFER ASH</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Severinia monophylla</td>
<td></td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Zanthoxyllum americanum</td>
<td>COMMON PRICKLYASH</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Zanthoxyllum clava-herculis</td>
<td>HERCULES'-CLUB</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Zanthoxyllum fagara</td>
<td>LIME PRICKLYASH, WILD LIME</td>
</tr>
<tr>
<td>Rutaceae</td>
<td>Zanthoxyllum flavum</td>
<td>WEST INDIAN SATINWOOD, YELLOWHEART, YELLOWWOOD</td>
</tr>
</tbody>
</table>
Another competing digitization priority are the specimens in loan requests. We image and database all specimens that are requested for loan then ask the requestor to further focus the request.
Development and enhancements of the database and front-end and curation of taxon-specific information (e.g., invasiveness) meant less time for digitization of the specimens (but also a product that we find particularly useful).
Outline

• Intro to the herbarium
• Our history of digitization
• The front-end
• Notes from Nature Project
• General observations
• **General recommendations**
General Recommendations

- Document the workflows and keep the documents up-to-date.
- Look over shoulders a lot at the beginning and provide feedback.
- Ask the digitizers how to streamline the process.
- Focus on making the specimens discoverable and useful for most research applications; it is not necessary to capture every specimen quality in the database initially.
- Consider adding “filed as” annotations to the specimens before photographing, if the name on the folder is not on the specimen label(s).
- Don’t worry if your names are not bleeding-edge current; synonymy relationships in databases can make them discoverable.
- Databasing from images makes remote work possible (e.g., from computer labs or Notes from Nature); this can be important when space is limited in the collection.
FSU’s Robert K. Godfrey Herbarium thanks the FL Fish and Wildlife Conservation Commission for their funding of a Task Assignment in the herbarium as well as NSF for the MorphBank grants (DBI 0446224 and 0851313), the Deep South Imaging grant (DBI 0646222), the Tall Timbers Digitization grant (DBI 0956343), and the iDigBio grant (EF 1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

To see the many people who have contributed to the productivity of the herbarium, visit http://herbarium.bio.fsu.edu/people.php.