GEOREFERENCING
TOOLS AND PRACTICES
Where does georeferencing fit into the digitization workflow?

- Traditional legacy approach: As complete records are entered, all data, including georeferencing data are determined and entered.

- New collections approach: Due to the prevalence of GPS devices, most modern collection data include georeferencing data which can be entered at the time specimen data area entered. Each collecting station becomes a collecting event that gets named, georeferenced, and entered. Collections at that station are all assigned to the same event and georeferenced as they are entered.
Where does georeferencing fit into the digitization workflow?

- New legacy approach 1: Georeferencing becomes an activity divorced from the initial entry of legacy data and dependent upon collaborative tools that allow records with similar localities to be parsed and pooled by software in a way that they are assigned georeferences in bulk.

- In bulk or individually based on known boundaries of counties, parks, preserves, sections, towns, counties, or other known areas with defined boundaries and known extents. In these cases, georeferencing is to reported accurate to the known extent. Such georeferencing can be built in to the data entry algorithm as a record is saved.
Georeferencing Data Elements

Required:
• Latitude (dd.dddd)
• Longitude (dd.dddd)
• Datum (coordinate system)

Desired:
• Precision
• Georeferencing method
• Georeferencing remarks
• Projection
Georeferencing Methods and Tools

ArcGIS

Google Earth/Maps

Biogeomancer

Geolocate
Drawbacks

- Slow
- Requires assembling and fiddling with map layers
- Potentially expensive
- Requires expert knowledge and familiarity with GIS and region
BioGeomancer:
Semi-automated Georeferencing Engine

http://bg.berkeley.edu/latest/

Developed by: John Wieczorek, Aaron Steele, Dave Neufeld, P. Bryan Heidorn, Robert Guralnick, Reed Beaman, Chris Frazier, Paul Flemons, Nelson Rios, Greg Hill, Youjun Guo
Collaborative Georeferencing

www.museum.tulane.edu/geolocate/

Online tutorial

Currently incorporated into Specify and Symbiota
Collaborative Georeferencing

The goal of this project is to provide a mechanism whereby groups of users can form communities to collaboratively georeference and verify a shared dataset. This collaborative georeferencing framework consists of two end-user components:

1. A client application for reviewing and editing community records. Currently there are 2 applications available for this task:
   - The GEOLocate desktop application
   - The GEOLocate web-based collaborative client (click link to try it out)

2. A web-based data management portal for creating and managing communities, their respective users and data sources

Shared community datasets created via the portal may consist of multiple underlying data sources from either live DiGIR providers and/or uploaded text files. Support for TAPIR providers is currently under development. Data are stored using the full Darwin Core 1.2 specification, but subsets and/or alternative schemas may be imported using the schema mapping interface. During ingestion, data items are automatically normalized, georeferenced and related to one another via a similarity index. This index is used to identify all records that appear to describe the same collection locality regardless of syntax. During coordinate verification, users have the option to re-classify those that were incorrectly related to one another.

Verification and correction of the computer generated geographic coordinates is accomplished using the GEOLocate desktop application. GEOLocate allows users to login to their communities, retrieve and visualize results, make any necessary corrections, provide additional comments, define errors as polygons, and save the results back to the shared database. The verified results of georeferencing can then be downloaded via the portal’s data management interface for re-import to the parent database.

To examine the gains in efficiency over traditional georeferencing 2100 randomly selected collecting events from the TUMNH fish collection were imported and georeferenced using the collaborative georeferencing framework. The TUMNH fish collection was georeferenced by hand in the mid to late 90’s and therefore provides a useful test bed for assessing the efficiency and accuracy of automated methodologies. Of the 2100 records, 30% were identified as being similar to other records and an additional 35% were duplicates leaving a total of 782 unique locations requiring correction, a 63% reduction in effort overall.

Video Tutorials
Welcome to the Tulane University Museum of Natural History’s Community Edition of GEOLocate. This site along with GEOLocate (version 3 and higher) and DIGIR provider software form the foundation of community-based georeferencing, whereby participants form communities and pool data to maximize efficiency of georeferencing.

Beta testing is now open, so if you are interested in using our software and services for collaborative georeferencing please follow the registration link below. After personal review of your registration information, you will be granted access to the system. Questions or comments may be directed to Nelson Ross.

Registration is free and gives you access to restricted areas of this portal, where you can join or create a community. Read more about us.

© 2012 Collaborative Georeferencing
Last updated February, 2012
Communities
<table>
<thead>
<tr>
<th>Data Source</th>
<th>Date Added</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duval1</td>
<td>Wednesday, June 27, 2012</td>
</tr>
<tr>
<td>Clay1</td>
<td>Wednesday, July 04, 2012</td>
</tr>
<tr>
<td>Clay2</td>
<td>Monday, July 09, 2012</td>
</tr>
<tr>
<td>Duval2</td>
<td>Sunday, July 29, 2012</td>
</tr>
<tr>
<td>Nassau1</td>
<td></td>
</tr>
</tbody>
</table>

Add a data source
## Community: Duval

### Data Sources

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Members</th>
<th>Member Email</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clay</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data source management operations**
- Add new community data source via DGIR.
- Add new community data source via CSV files.

**Click on an item's header to expand/collapse its content.**

- **Community-wide**
  - Duval
  - Clay

**Secondary name:**
- Clay

**Path:**
- CSV

**Type:**
- CSV

**Owner:**
- CSV

**Number of records:**
- CSV

**Status:**
- CSV

**Task:**
- CSV

**Records statistic:**
- Total corrected community-wide 750
- [View corrections]

**Total corrected in this data source:**
- [View corrections]
Edit:
- Match Water Body
- Do Uncertainty
- Detect Hwy/River Crossing
- Restrict to lowest Adm. Unit
- Displace Polygon
- Language: English

Privacy setting:
- private

Edit:
- public: any user can join your community.
- restricted: users send you requests to join your community.
- private: only users you invite can join your community.

Description:

Edit:

Your search query:
- You are restricted to Nassau.

Edit:
To edit a user's search query, use the “Define users working dataset” interface under the “Members” tab.

Use this easy signup link to send a auto-join invite to users:

Enable hyperlinks on data source records:
- base link format: http://domain.com/?<key>= where <key> is a parameter key of your choice.

DISABLE EASY SIGNUP LINK
REMOVE SELECTED HYPERLINK
ADD SPECIMEN HYPERLINK
Collaborative Georeferencing

The goal of this project is to provide a mechanism whereby groups of users can form communities to collaborate and verify a shared dataset. This collaborative georeferencing framework consists of two end-user components:

1. A client application for reviewing and editing community records. Currently there are 2 applications available for this task:
   - The GEOLocate desktop application
   - The GEOLocate web-based collaborative client (click link to try it out)

Web-based data management portal for creating and managing communities, their respective users and data sources

Shared community datasets created via the portal may consist of multiple underlying data sources from either live DigiGIR providers and/or uploaded text files. Support for TAPIR providers is currently under development. Data are stored using the full Darwin Core 1.2 specification, but subsets and/or alternatives schemas may be imported using the schema mapping interface. During import, data items are automatically normalized, georeferenced and related to one another via a similarity index. This index is used to identify all records that appear to describe the same collection locality regardless of syntax. During coordinate verification, users have the option to re-classify records that were incorrectly related to one another.

Verification and correction of the computer generated geographic coordinates is accomplished using the GEOLocate desktop application. GEOLocate allows users to login to their communities, retrieve and visualize results, make any necessary corrections, provide additional comments, define errors as polygons, and save the results back to the shared dataset. The verified results of georeferencing can then be downloaded via the portal’s data management interface for re-import to the parent database.

To examine the gains in efficiency over traditional georeferencing 2100 randomly selected collecting events from the TUMNH fish collection were imported and georeferenced using the collaborative georeferencing framework. The TUMNH fish collection was georeferenced by hand in the mid to late 90’s and therefore provides a useful test bed for assessing the efficiency and accuracy of automated methodologies. Of the 2100 records, 30% were identified as being similar to other records and an additional 35% were duplicates leaving a total of 782 unique locations requiring correction, a 63% reduction in effort overall.

Video Tutorials
Login with same credentials as the management portal

www.museum.tulane.edu/geolocate/web/webcomgeoref.aspx
Click Continue
Locality string

Mark placed roughly at the point of locality string…..or not!

Shaded circle=radius of uncertainty.

Click the specimen id to access the original record.
Specimen Details

Species: Scleria triglomerata
Collection Date: June 2, 1968
Bar Code ID: 000027481
Collector: R.K. Godfrey, William Lindsey
Collector's Identifier: 56998
Flower Buds Present?: True
Flowers Present?: True
Fruit Present?: True
Country: United States
State or Province: Florida
County or Parish: Nassau
Fips Code: 12089
Nearest Named Place: Crawford
Verbatim Directions to Locality: Vicinity of Crawford.
Habitat: Pine flatwoods
Non digital information?: N

Identification Records
Identifier's Name: R. K. G.
Formal Identification Date: 1968-06-02
Lowest Rank Identified: Species
Identification History: Scleria triglomerata Michx.
Determinant: n
*Scleria triglomerata Nutt.*

**PLANTS OF FLORIDA**

**HERBARIUM OF FLORIDA STATE UNIVERSITY**

**TALLAHASSEE**

**COUNTY:** BASSAU  
June 7, 1980  

**Scleria triglomerata Nutt.**  

Fine flatwood, vicinity of Lynnwood.

Collected by M. W. Soffrey & William Lindsey,  
No. 50026,  
Det. T. D. Estes

Collected by L. W. Soffrey II & William Lindsey,  
No. 50063,  
Det. T. D. Estes

Collected by W. D. Soffrey II & William Lindsey,  
No. 50064,  
Det. T. D. Estes

Collected by W. D. Soffrey II & William Lindsey,  
No. 50065,  
Det. T. D. Estes

Collected by W. D. Soffrey II & William Lindsey,  
No. 50066,  
Det. T. D. Estes

Collected by W. D. Soffrey II & William Lindsey,  
No. 50067,  
Det. T. D. Estes

**SEARCH FLORIDA STATE**

FSU Biology  

View Image  
View JPEG  
Download JPEG (1.35 MB)  
Download TIFF (30.83 MB)
Uncertainty
Edit uncertainty
Determining uncertainty can be subjective and is not always easy. Effectively, georeferencers should strive to use physical features to draw their circle or polygon. Such features might include roads, city or park boundaries, landscape features (lakes, bogs, habitat signatures, etc.), and other known demarcations. Attend additional trainings and establish institutionally specific policies and protocols, which may include manifest estimates of confidence.
## Community: Duval

### Data Sources

<table>
<thead>
<tr>
<th>Data source</th>
<th>Date added</th>
<th>Secondary name</th>
<th>Path</th>
<th>Type</th>
<th>Owner</th>
<th>Number of records</th>
<th>Status</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>community-wide</td>
<td>Wednesday, June 27, 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duval1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City1</td>
<td>Sunday, July 29, 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City2</td>
<td>Monday, July 9, 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duval2</td>
<td>Monday, July 9, 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Records statistics

- 762 total corrected specimen records
- 818 total locality records
- 52 corrected specimen records

[View corrections]
GOAL: Import corrections into collections database.

Download corrections

Polygon data