



# Mid-Atlantic Megalopolis Georeferencing Guidelines

Version 1

April 2019

Morris Arboretum of University of Pennsylvania  
Philadelphia, Pennsylvania

Written by Michelle Mancini, Anne Barber, Timothy A. Block and C. Skema

Logo design by Riley Design & Illustration  
© 2016

Project funded by  
the Advancing Digitization of Biological Collections Program  
of the National Science Foundation  
Awards 1601697, 1600981, 1601393, 1601429, 1600976,  
1601101, 1601503, 1743744, 1702441, and 1801022.

## Contents

1. Introduction .....	4
1.1. The Salient Points .....	4
1.2. Terminology.....	4
<i>Table 1.</i> Definitions of georeferencing terminology and tools. ....	4
1.3. Workflow & Training.....	5
<i>Figure 1.</i> General georeferencing workflow. ....	6
2. Clean, Sort & Categorize Records Before Georeferencing .....	7
2.1. Processing Status Categories.....	7
2.2. Clean with Geography Cleaning Tools .....	7
2.3. Sort Out Problematic Records, “Do Not Georeference” – Phase 1 of 2.....	7
2.4. Sort & Evaluate Records with Coordinates.....	8
2.4.1. Sort Out Label Coordinates from GPS Devices.....	8
2.4.2. Sort Out Label Coordinates from Other Methods.....	8
3. Georeferencing .....	9
3.1. Data Fields .....	9
<i>Figure 2.</i> Data fields relevant for georeferencing .....	9
<i>Table 2.</i> Explanation of data fields to be used in georeferencing.....	10
3.2. Georeferencing with GeoLocate.....	11
<i>Figure 3.</i> An overview of the use of GeoLocate within Symbiota.....	12
<i>Box:</i> A Word On Uncertainty.....	13
<i>Box:</i> Error Polygons: To Use or Not to Use? .....	14
<i>Box:</i> Error Polygons in GeoLocate: Instructions for Use .....	15
3.3. Sort Out Problematic Localities, “Do Not Georeference” – Phase 2 of 2.....	16
3.4. Manual Georeferencing.....	17
3.5. CoGe: A Collaborative Georeferencing Tool.....	17
4. Review & Clean Data .....	18
5. Citations & Resources .....	20
5.1. Citations.....	20
5.2. Resources about the Practice of Georeferencing.....	20
5.3. Resources for Use in Georeferencing .....	20
<i>Table 3.</i> Resources for use in georeferencing.....	20

## 1. Introduction

The purpose of this document is to standardize our approach to georeferencing as a collaborative Thematic Collections Network. Georeferencing is the process of assigning geographic coordinates to specimens as accurately as possible based on whatever locality information is given on the specimen label. The primary tool we will use for this process is GeoLocate (Rios, 2018) which is integrated into Symbiota (Symbiota Working Group, 2018) in the MAM portal. These guidelines assume a working familiarity with the Symbiota software in the MAM portal. For anyone lacking that, please see the MAM Manual (Skema & Barber, 2018).

### 1.1. The Salient Points

- For the MAM Project, we recommend georeferencing records for which, at a minimum, the county is known. Typically, that means do not georeference records for which the only locality information available is at the state level or above. It is generally not useful for research efforts to assign a “centroid” point at such a coarse scale as state (or above).
- Do not assign finer granularity of locality to a specimen than exists on the label unless you have personal knowledge of the specimen (e.g., you were on the collecting trip), access to field notes, exceptional familiarity with collectors and their habits, or access to other information that allows you to correctly refine the geography of the locality on the label.
- Realize that not all records can be georeferenced! Various problems with locality descriptions, as detailed in Section 3.3, can prevent the correct assignment of coordinates. The percentage of records that cannot be georeferenced will vary by collection.

### 1.2. Terminology

Table 1 provides definitions for the terminology used throughout this document and describes some of the main tools in Symbiota to be employed.

**Table 1.** Definitions of georeferencing terminology and tools.

Term (Synonyms)	Definition	Notes
Batch Editor Tool	The tool in Symbiota that allows editing of a selected data field across multiple records at once.	Instructions for use can be found in the MAM Digitization and Training Manual, section 8.1.5. <i>Batch Editing</i> .
Batch Georeferencing Tool	The tool in Symbiota that allows georeferencing of multiple records with the same or similar localities.	Instructions for use can be found in the MAM Digitization Standards and Training Manual, section 10 <i>Georeferencing</i> .
GeoLocate in Occurrence Editor	The tool in Symbiota that allows georeferencing of records one-by-one.	Instructions for use can be found in the MAM Digitization Standards and Training Manual, section 10 <i>Georeferencing</i> .
Heading	The direction of displacement from a particular place (such as a town).	Typically given in conjunction with an offset. e.g., 0.5 miles <b>W of Perryville</b>
Offset	The distance of displacement from a particular place (such as a town).	Typically given in conjunction with a heading. e.g., <b>0.5 miles W of Perryville</b>

Term (Synonyms)	Definition	Notes
Point	The intersection of a single latitude value and a single longitude value.	e.g., GeoLocate automatically detects the known decimal latitude and longitude point of the center of “Philadelphia” as 40.009376, -75.133346
Point-radius method	The process of georeferencing using a single point and an uncertainty error as a circle that extends a particular radius from that point.	This is currently the standard practice used in georeferencing. See box “A Word On Uncertainty” for more details.
Polygon (Footprint WKT) (Footprint Polygon) (Error Polygon)	The shape drawn to define a known locality, consisting of multiple decimal latitude and longitude vertices.	e.g., GeoLocate detects the shape of “Philadelphia” based on known decimal latitude and longitude vertices, if you select the Do Error Polygon setting under Options.
Uncertainty (Uncertainty circle) (Error circle) (Error radius)	The radius in meters measured outward from a single point, representing the calculation of the potential range of error for that position.	See box “A Word On Uncertainty” for more details. e.g., GeoLocate automatically calculates uncertainty for “Philadelphia” as a radius of 22,256 m.

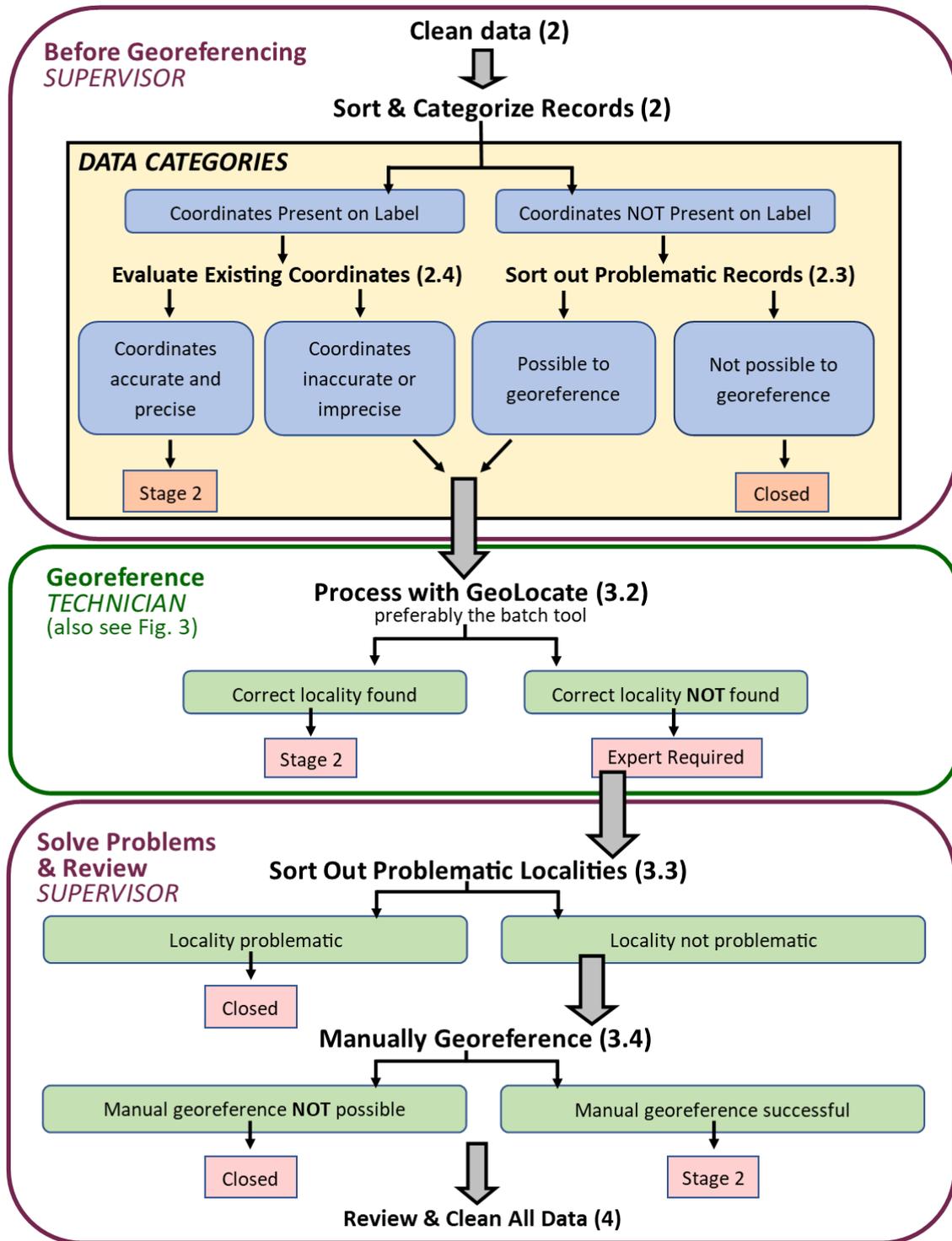
### 1.3. Workflow & Training

Georeferencing should begin after all records (or as many as possible) have been transcribed. This allows you to take advantage of the time-saving functions of batch georeferencing in Symbiota. Batch georeferencing alleviates the need to georeference the same locality multiple times.

It is very important for anyone doing georeferencing to have, or quickly develop, a detailed knowledge of the geography of the areas covered by a collection. To facilitate this, consider assigning individual students or technicians to limited geographic areas (e.g., a county, preserve, or state, if need be), rather than having each person georeferencing all over the Mid-Atlantic. Furthermore, as compared to transcription, georeferencing is harder to review afterwards for correctness, so thorough training and follow-up with technicians at the beginning is essential.

Figure 1 shows a model for the general workflow recommended for georeferencing in the MAM Project. The following sections will provide further explanation but the general idea is to do some cleaning and sorting of records prior to georeferencing, then begin with the Georeferencing Batch Tool, first targeting the records with the simplest locality descriptions that are correctly detected by GeoLocate. This step will allow technicians to practice using GeoLocate, learn to properly review georeferencing results from GeoLocate before saving them back into the portal, and build confidence in the process. Also, this step likely will complete the digitization of a substantial portion of your collection! Unavoidably, there will be some “difficult” records that will require more time to georeference. We recommend setting these aside in the processing status category Expert Required, then return to them later. The experience of georeferencing the simpler records first, which are typically the bulk of the collection, may help clarify the issues with the more difficult records and, at the least, will not hamper progress. (This is true whether or not the work is split between a supervisor and a technician, or if it is all being done by a single person.) Finally, some of these records may need to be manually georeferenced or closed if it is not possible to georeference them.

**Figure 1.** General georeferencing workflow. Numbers in parentheses refer to sections in the text. Pink boxes signify a processing status for a record in the MAM portal. Expert Required is a holding place for records that need a supervisor’s attention after a technician has unsuccessfully attempted to georeference them, while Stage 2 and Closed are an endpoint for the georeferencing workflow for that sheet, with the exception of a final review.



## 2. Clean, Sort & Categorize Records Before Georeferencing

### 2.1. Processing Status Categories

The correct use of processing status categories is critical to organization and progress in any digitization effort in MAM, including georeferencing. The processing status categories employed in this document follow those used in the MAM manual<sup>1</sup> with one improvement, as follows. During the georeferencing process, it is recommended to employ the processing status “Closed” as a place to corral records that cannot be georeferenced, as separate from the rest. In this scenario, all records that have been imaged, transcribed, reviewed, and georeferenced successfully would be in Stage 2, while all records that have been imaged, transcribed, and reviewed but are unable to be georeferenced would be in Closed. Having a separate place to put such records keeps technicians from coming across them repeatedly during georeferencing, and helps keep the workflow clear in the database. It also enables quick retrieval of specimen numbers in more meaningful categories for writing progress reports.

### 2.2. Clean with Geography Cleaning Tools

It is best to start georeferencing with clean data. The Data Cleaning Tools should only be used once you are confident that records have been transcribed accurately. In the Administration Control Panel, under the General Maintenance Tasks heading, click on Data Cleaning Tools. Scroll down to the Political Geography section and click on Geographic Distributions. In the lists provided, you can drill down to the county level starting at the country level. This simple tool provides you with an easy way of finding possible mistaken entries. For example, “United Arab Emirates” often gets picked from a dropdown list instead of “United States” or technicians type “USA” instead of the standard “United States.” You can find those mistakes here and then fix them in the Batch Editor Tool. Next, go back to the Data Cleaning Tools and click on Geography Cleaning Tools. An analysis will return any values in the country, state, and county fields that don’t match the central geography authority and give you an opportunity to fix them. It will also return any null values and give you an opportunity to autofill them.

### 2.3. Sort Out Problematic Records, “Do Not Georeference” – Phase 1 of 2

Before you begin the georeferencing process, sort out records that are lacking sufficient locality information to be georeferenced. For MAM, it is recommended to only georeference records that have locality descriptions at the county level or below. Search for records that do not match this criterion by selecting the following in Edit Existing Occurrence Records:

Processing Status: Stage 1  
Custom Field 1: County IS NULL  
Custom Field 2: Locality IS NULL (some institutions may have “no locality” or similar in locality field, refine search as appropriate)

Verify that detailed locality information is indeed missing by checking the data against the labels. Then, add a brief explanation in the Georeference Remarks field of why the record was closed and change the processing status to Closed.

---

<sup>1</sup> A summary of the use of categories for Processing Status in MAM portal:

Unprocessed = imaged with skeletal data

Pending Review/Expert Required/Reviewed = imaged + in process of transcription/review

Stage 1 = imaged + transcribed + reviewed

Stage 2 = imaged + transcribed + reviewed + georeferenced; Stage 3 = record problematic and needs more investigation

## 2.4. Sort & Evaluate Records with Coordinates

Next sort out records that have coordinates present on the label. These should have been entered during transcription in the Latitude and Longitude fields, with “coordinates recorded at time of collection” entered in the Georeference Remarks field. The Verbatim Coordinates field may also be filled in if the coordinates on the label were not in full decimal format. Before beginning the sorting process, it may be useful to check that the Decimal Latitude and Decimal Longitude fields are not blank in instances when the Verbatim Coordinates field has been transcribed to ensure that Verbatim Coordinates were converted properly to decimal format. Refer to the section “Transcription & Review” in the MAM manual for instructions on accurately transcribing and reviewing this information from the label.

### 2.4.1. Sort Out Label Coordinates from GPS Devices

Search for records with coordinates collected later than 1990, as these coordinates were likely recorded with GPS units.

*Note:* Prior to 1990, the precision of the methods used to record coordinates may not have been sufficient to be accepted as-is and those records may need to be georeferenced. (See section 3.4.2 below).

Search for records with post-1990 coordinates by selecting the following criteria in Edit Existing Occurrence Records:

Custom Field 1: Decimal Latitude IS NOT NULL

Custom Field 2: Date GREATER THAN “1990-00-00”

Verify the database coordinates against the label coordinates, as needed, and check that the coordinates map to the position described in the locality. (Numbers can easily be transposed in transcribing coordinates for field notes or labels!) Confirm that “coordinates recorded at time of collection” is present in the Georeference Remarks field. Use the Batch Editor Tool to autofill the information below in the respective fields:

Georeference Protocol: “MAM Georeferencing Policies and Procedures (2019)”

Uncertainty\*: “10” (or enter uncertainty from label, if present)

Processing Status: Stage 2

\*In the absence of other information, an uncertainty of 10 m is recommended because that is a reasonable estimate of the accuracy for most GPS units at typical geodetic datum settings under normal field conditions.

### 2.4.2. Sort Out Label Coordinates from Other Methods

Search for records with coordinates collected prior to 1990, as these coordinates were likely to have been recorded using another method, like triangulation on paper maps. Search for them by selecting the following criteria in Edit Existing Occurrence Records:

Custom Field 1: Decimal Latitude IS NOT NULL

Custom Field 2: Date LESS THAN “1990-00-00”

Coordinates recorded on the label to degrees, minutes, and seconds are generally acceptable. Depending on the level of review of records completed after transcription, you may want to verify the coordinates against the label. Use the Batch Editor Tool to autofill the information below in their respective fields:

Georeference Remarks: “coordinates recorded at time of collection”, *if not already filled in during transcription*

Georeference Protocol: "MAM Georeferencing Policies and Procedures (2019)"

Uncertainty:

- Enter uncertainty from label, if present
- If uncertainty is not present on label, enter "30 m" for coordinates recorded to seconds, if locality is in the Mid-Atlantic

Processing Status: Stage 2

Coordinates recorded on the label to degrees and minutes only are generally unacceptable. Such specimens should be georeferenced. For the records that need to be georeferenced, remove data from all of the georeferencing fields except Verbatim Coordinates, where coordinates should have been entered exactly as they appear on the label. Include a note in "Georeference Remarks" to indicate why the record needs to be georeferenced (e.g., imprecise coordinates on label). Keep the Processing Status for these records as Stage 1.

### 3. Georeferencing

#### 3.1. Data Fields

Georeferencing data fields are shown in Figure 2 and listed with explanations on their usage in Table 2. Fields that are required are marked with an asterisk. Note that the "Georeferenced By" and "Georeference Sources" fields do not need to be filled in if the coordinates were recorded at time of collection.

Realize that there are two ways to georeference within Symbiota: within the Occurrence Editor, clicking on the GeoLocate tool for each specimen one by one, or from within the Georeferencing Batch Tool. These two tools behave differently from one another in that each will autopopulate different fields for each record. These behaviors are described in Table 2.

**Figure 2.** Data fields relevant for georeferencing in the Occurrence Editor in the MAM Portal. Letters correspond to field explanations in Table 2 below. Required fields are marked with an asterisk.

The screenshot shows a form with the following fields and labels:

- Latitude **a\***
- Longitude **b\***
- Uncertainty **c\***
- Datum **d**
- Verbatim Coordinates **e**
- Elevation in Meters
- Verbatim Elevation
- Depth in Meters
- Verbatim Depth
- Georeferenced By **f\***
- Georeference Sources **g\***
- Georeference Remarks **h**
- Georeference Protocol **i\***
- Georef Verification Status **j**
- Footprint Polygon **k**

**Table 2.** Explanation of data fields to be used in georeferencing in the MAM Portal. Required fields are marked with an asterisk.

Field	Description	Example	Autopopulates when using Georeferencing tools?
<b>a) Latitude*</b>	The geographic latitude in decimal degrees	44.32611	Yes, both in Georeferencing Batch Tool and in GeoLocate in Occurrence Editor
<b>b) Longitude*</b>	The geographic longitude in decimal degrees	-74.130833	Yes, both in Georeferencing Batch Tool and in GeoLocate in Occurrence Editor
<b>c) Uncertainty*</b>	Uncertainty in meters, measured as a radius	1000	Yes, both in Georeferencing Batch Tool and in GeoLocate in Occurrence Editor
<b>d) Datum</b>	Geographic model used	WGS84	Only if using GeoLocate in the Occurrence Editor
<p><i>Both Google Maps and GeoLocate deliver coordinates in WGS84. If coordinates are present and the datum is referenced on the label, you will need to manually enter that datum. If that datum is not WGS84, you could convert the coordinates<sup>2</sup> from the datum referenced on the label to WGS84 to standardize the data in your collection. When using GeoLocate within the Georeferencing Batch Tool, you will need to enter WGS84 in the Datum field just before saving the data. Refer to the MAM Digitization and Training Manual for instructions on entering data into fields within the Georeferencing Batch Tool.</i></p>			
<b>e) Verbatim Coordinates</b>	Use this field to enter any coordinates present on label that are not in decimal format.	44° 19' 34" N, 74° 7' 51" W	No, must enter manually
<p><i>When using this field, also complete field h) Georeferencing Remarks. Instructions for entering this information in the Occurrence Editor is described in the MAM Digitization and Training Manual section 8.1.2. Quick Guide to Occurrence Data in Record Editor</i></p>			
<b>f) Georeferenced By*</b>	Username of user who did the georeferencing	abarber	Only in the Georeferencing Batch Tool, must enter manually otherwise
<b>g) Georeference Sources*</b>	Sources used to obtain coordinates	GeoLocate Google Earth other resources	Yes, both in Georeferencing Batch Tool and in GeoLocate in Occurrence Editor
<b>h) Georeference Remarks *Required if coordinates are present on label</b>	Use to note specimens for which coordinates were present on label OR Notes about the georeferencing	“coordinates recorded at time of collection” OR “cannot georeference because...”	No, must enter manually as needed

<sup>2</sup> A good online tool for coordinate conversion exists at <https://tagis.dep.wv.gov/convert/>.

Field	Description	Example	Autopopulates when using Georeferencing tools?
<i>If coordinates are present on the label, enter “coordinates recorded at time of collection” in Georeference Remarks. It is useful to standardize the Georeferencing Remarks and Notes as much as possible within your collection.</i>			
<b>i) Georeferencing Protocol*</b>	Protocol used to georeference	MAM Georeferencing Policies and Procedures (2019)	No, must enter manually
<i>Instructions for entering the Georeferencing Protocol in batch in the Batch Editor Tool is described in the MAM Digitization and Training Manual section 8.1.5. Batch Editing.</i>			
<b>j) Georef. Verification Status</b>	Indicates confidence level	reviewed – high confidence	Only if working in the Georeferencing Batch Tool
<i>Georeferencing Verification Status auto-populates in the Georeferencing Batch Tool as “reviewed – high confidence.” The assumption is that the information has been reviewed and the confidence is high. You can leave this at the default value, adjust it, or remove altogether.</i>			
<b>k) Footprint Polygon</b>	Polygon drawn between vertices of decimal latitude and longitude	42.4665920229,- 76.5125549066, 42.4666270229,- 76.5105929066, 42.4668370229,- 76.501542906...	If you have turned on Do Error Polygon in GeoLocate, and a polygon is detected, it will autopopulate unless you select “clear polygon.” For details on the use of polygon data, read the Error Polygons box.

### 3.2. Georeferencing with GeoLocate

GeoLocate is the preferred tool for georeferencing specimen records in the MAM Project. GeoLocate interprets locality text strings and assigns coordinates to them based on its geographic database. GeoLocate also provides a measure of uncertainty using the georeferencing calculator developed by MaNIS, HerpNet, and ORNIS (<http://manisnet.org/gci2.html>). These coordinates (latitude + longitude) and the uncertainty (in meters) are the data that should be imported from GeoLocate back into Symbiota for all records. Also, you may or may not want to import/create polygon data, depending on the locality. See the Error Polygons box for more details on this topic.

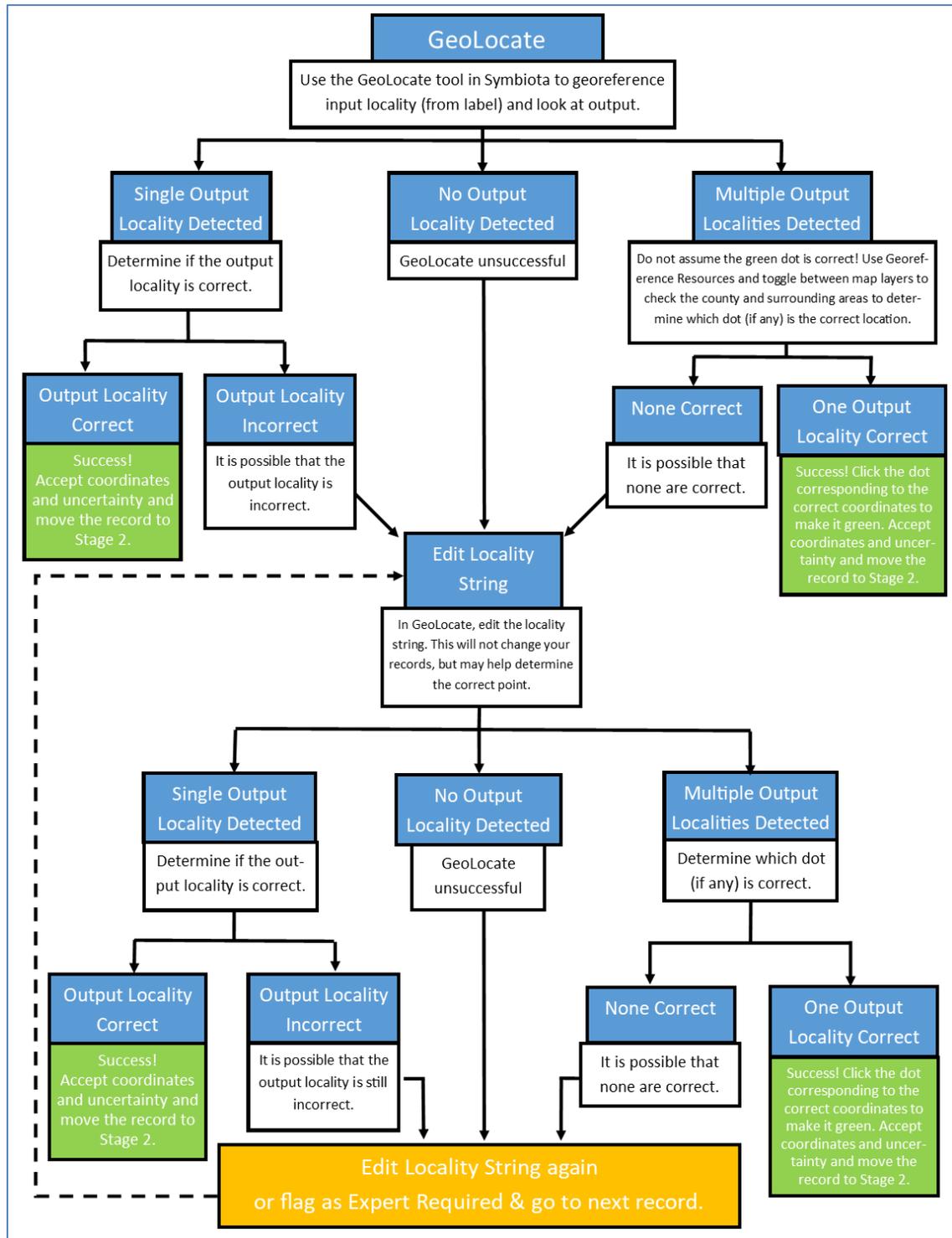
See the MAM Manual Section 10 Georeferencing for step-by-step instructions on how to use GeoLocate in the portal.

Using GeoLocate through the Georeferencing Batch Tool in Symbiota is the most efficient way to georeference in the MAM Project, so start your work in the batch tool. The decision tree in Fig. 3 can help to guide you through the GeoLocate process.

**Results from GeoLocate should never be accepted at face value without checking that the output locality is correct!** For any point, actively evaluate if the point placement is correct by checking against higher level geographic information (e.g., is it in the right county?) and verifying that it meets the locality qualifiers. Use other resources, online and on paper (e.g., quad maps), as need be. (See Section 5.3 for a partial list of potential resources.)

If more than one point is returned by GeoLocate, decipher the parsing pattern to better understand what GeoLocate is working with to produce these points. Understanding what GeoLocate is parsing on can also point out mistakes in the results, for both single and multiple point output. This information can

**Figure 3.** An overview of the use of GeoLocate within Symbiota. See this document for general procedure and the MAM Manual for step-by-step instructions.



## A Word On Uncertainty

A single designated point (i.e., the intersection of a particular latitude and longitude) plus the uncertainty drawn as a circle extending from that point are the two components of the “point-radius” method of georeferencing, which is the current standard procedure for georeferencing. But what is uncertainty?

Uncertainty (also known as uncertainty circle, error circle, or error radius) is a calculation of the potential error for the position of the point. It has been said that it is the “sum of what we know about what we don’t know.” The calculation of uncertainty is fairly standardized and depends on various characteristics of the locality statement, such as the following:

- extent of the specific named place within the locality string
- imprecision in distance measurements
- imprecision in coordinate measurements
- imprecision in direction measurements
- map scale

For records without offsets, the type of feature being described also matters, hence why you will likely see some commonly repeated uncertainty numbers generated by GeoLocate for similar types of features/places (e.g., 3036 m for an unbounded named place). In our workflow, GeoLocate will calculate the uncertainty for you, unless you choose to modify the radius of the uncertainty circle yourself.

Given that uncertainty is often understood as an indication of the quality level of the georeferenced data for use in research, there is a widespread appreciation that it is better to overestimate the error rather than underestimate it. However, for certain features or localities, there may be a desire to refine this uncertainty with a polygon. See the box “Error Polygons: To Use or Not To Use?” for more details.

For more information on uncertainty see the section “Calculating Uncertainties” in Chapman & Wieczorek (2006).

then be used to try editing the locality string to improve georeferencing performance, if needed. (See “Parse pattern” in the GeoLocate Quick Guide in the MAM Manual for more details on finding parsing information.)

Linear features, such as streams or ridges, can be problematic in GeoLocate. Often the point locality returned by GeoLocate for these types of features will be the calculated centroid, or a point placed at one end or the other of the feature.

*Example 1:* If the label reads “Delaware River, Bucks County, Pennsylvania” GeoLocate will return two points, neither of which is in Bucks County. One is at the “mouth” of the Delaware River on the boundary between Delaware and New Jersey, and the other at the “centroid” of the Delaware River in Pike County, PA between Bushkill and Dingman’s Ferry.

*Example 2:* A label that reads “Blue Mountain, Lehigh County, Pennsylvania” will result in a single point placed near Roxbury in Franklin County, PA in spite of the fact that the ridge known as “Blue Mountain” in Pennsylvania stretches from the Maryland boundary in Franklin County to the New Jersey boundary along the Delaware River.

In both of the above examples, changing the county designation has no effect on the results from GeoLocate. Clearly, neither of these are correct and these specimens need to be manually georeferenced.

## Error Polygons: To Use or Not to Use?

*See instructions in the box "Error Polygons in GeoLocate: Instructions for Use" for how to get/accept/draw polygons in GeoLocate.*

Generally, in MAM we recommend only using polygon data for a locality that is a simple bounded named place with no offset, for which a polygon already exists in GeoLocate.

*Example:* The specimen locality is simply "Philadelphia." It is reasonable to accept the point given for Philadelphia, which is typically the standard gazeteer point for the place name, and the uncertainty circle as given. However, the potential boundaries for where this specimen was collected could be further refined by accepting the error polygon generated by GeoLocate.

If you have records which have a simple bounded named place with no offset for which NO polygon exists in GeoLocate, it is up to the collection/digitization manager as to whether one should spend the time drawing the polygon (or retrieving a known polygon from an outside data source) or simply accept the provided point + uncertainty circle as sufficient.

There may be some extraordinary records for which the locality is too general to allow non-arbitrary point georeferencing, but specific enough to allow for the creation of an informative polygon. In these instances, it may be worth creating a polygon.

*Example:* The specimen locality is "Wissahickon Valley Park, between Northwestern Avenue and Bells Mill Road." Although you do not know the exact point where the specimen was collected, you do know a well-defined space in which it was collected. In this instance, one could draw a polygon bounded by Northwestern Avenue, two park boundaries, and Bells Mill Road.

In some georeferencing communities, there is also a practice of drawing polygons (or retrieving a known polygon from an outside data source) for linear (e.g., rivers, creeks, rail lines) or odd-shaped features (e.g., lakes, preserves) where the uncertainty circle is evaluated as simply too much of an overestimate.

In the first example, the polygon already exists and merely has to be imported into the digital collection, hence our recommendation for the consistent use of this practice in MAM. In all but the first example, each institution has to weigh the time invested in creating/importing the polygon against the information gained. Some think that the relatively rough point-radius method for georeferencing will eventually be replaced by an exclusive use of polygons, however at present polygons are not readily available for many features/places and they can be time-consuming to draw/import.

In thinking about future research, the examples above might be better served, at least in some studies, by the exclusive use of the polygon alone as the georeferenced "point" (e.g., averaging relevant geographic data across all points included in the polygon), but at this time it is standard practice for the georeferenced data to still include a single chosen point (even if arbitrary) and an uncertainty circle along with the polygon. Of course, adding the footprint polygon to the data, where appropriate, does make it available to researchers who seek to use it in lieu of a single point. Also, where appropriate, leaving a note in Georeference Remarks on the arbitrary nature of the point serves as a warning to potential researchers that it maybe be best to use the polygon in lieu of the single point.

## Error Polygons in GeoLocate: Instructions for Use

GeoLocate's default options do not show the error polygon so you will need to turn on polygon detection for localities for which you are seeking a polygon (e.g., a simple bounded named place with no offset).

To turn on polygon detection:

1. Within the GeoLocate application, select the "Options" tab.
2. Check the box next to "Do Error Polygon."
3. Click "Georeference" to re-georeference the current locality string including the error polygon. If no polygon is available, you will see "Unavailable" come up in the returned data box (lower right of screen, above Save To Your Application button).

Do not select "Do Error Polygon" if you are working with offsets or locality strings with more detail. Alternatively, you can click the "Clear Polygon" button to remove the data and not take it into your application, or manually delete the polygon data from Footprint Polygon field once it is saved back to Symbiota, if it is not wanted.

You may want to draw a polygon for a particular record. (See box "Error Polygons: To Use or Not to Use?" for a discussion on when to draw/import polygons.)

To draw a polygon using the GeoLocate tool in Symbiota:

1. In GeoLocate tool (both in batch georeferencing and one-by-one in Occurrence Editor), click on "Draw polygon."
2. Click on the first point of the outline of your polygon, then continue to click on points on the map until you have outlined the shape you want.
3. Double click to close the polygon (it will connect your last point to your first point by the shortest path possible).
4. In some instances, you may want to retain the point placement and standard uncertainty measurement given, while in others you may want the point + uncertainty circle to mirror the polygon as closely as possible. In the latter instance, drag the dot to an arbitrary point in the polygon, click on the dot and then click on "Resize uncertainty to polygon."
5. The polygon vertices will be listed in the returned data box. Click "Save To Your Application" to save them to Symbiota.
6. Make a note about drawing the polygon (and, if appropriate, the arbitrary placement of the point) in Georeference Remarks. For example, one could enter "locality description drawn as polygon, arbitrary point placed within polygon, uncertainty circle resized to fit polygon" in Georeference Remarks.

If you have progressed through the decision tree in Fig. 3 and GeoLocate still fails to detect the correct locality, mark the record as Expert Required<sup>3</sup>, write a concise remark in Notes as to what the problem is,

---

<sup>3</sup> To change processing status in the Batch Georeferencing Tool, one must enter values into the Decimal Latitude and Longitude fields. Simply enter a placeholder of "1" in each, but remember that those records will not come up when you search for Decimal Latitude/Longitude as NULL to find records that still need to be georeferenced. They can be found again by searching for Processing Status Expert Required (or, if there are additional records in Expert Required, single them out by a search on Custom Field 1: Decimal Latitude EQUALS 1). [If you are concerned about entering erroneous lat/long, even temporarily, realize these specimens will map to the Gulf of Guinea, so will stand out as an obvious error to both you and potential researchers.]

and move on to the next record. Once the simple records have been georeferenced and the harder records are corralled in Expert Required, return to the latter and proceed to the next section of these guidelines.

### 3.3. Sort Out Problematic Localities, “Do Not Georeference” – Phase 2 of 2

If you are unable to determine coordinates after attempting to use GeoLocate, the record(s) should be moved to either Expert Required for further review later or Closed if the locality cannot be georeferenced. Geographic certainty should not be applied to the record if that certainty cannot be obtained from the original data (potentially including auxiliary material, such as field notebooks and other collections by that collector on that day). Always include an explanation in Notes of why the record was moved to Expert Required. When moving a record to Closed, always include “not georeferenced” followed by a concise explanation of why it could not be georeferenced in Georeference Remarks. Localities that are not able to be georeferenced typically fall into one of these categories:

1. *Locality Ambiguous* – If the locality is ambiguous, or if there are multiple places with a similar name and the label does not provide a clear indication of which locality is correct, geographic certainty cannot be applied. It may be helpful to check the label first for additional information. In some instances, it may be possible to determine the correct locality through field notes, or sure knowledge of the collector’s survey areas. If you are unable to determine a correct locality with confidence, change the processing status to “Closed” and include a concise explanation in “Georeferencing Remarks” as to why it was closed.  
*Example:* The locality reads “Chestnut Hill, Pennsylvania.” There is more than one locality in Pennsylvania named Chestnut Hill. If the label does not contain county information or other identifiers that would help determine which locality to choose, the locality cannot be georeferenced.
2. *Locality General or Broad* – If the locality description is overly general, a georeferenced point for such a record may be less useful for downstream scientific research use. It is recommended for MAM that any locality at the county level or below should be georeferenced, whenever possible, while anything at the state level or above should not be. Many of these records may have already been sorted out earlier (see Section 2.3), though there may still be some records with general or overly broad information in the locality field.  
*Example:* Localities such as “Pennsylvania” or “Eastern US” are overly general. MAM policy does not recommend entering a state, or regional, centroid above the county level.
3. *Locality Contains Uncertainty* – If the locality description was originally recorded with uncertainty, then the record cannot be georeferenced and should be moved to “Closed” with a concise explanation in “Georeferencing Remarks.”  
*Example:* Localities such as “Bucks County?” or “presumably Longwood Gardens” contain uncertainty as denoted by “?” and “presumably” and should not be georeferenced.
4. *Locality Contains Contradiction* – If the locality description contradicts itself, the most likely assumption is that it was recorded incorrectly either at the time of collection or label creation and therefore geographic certainty should not be applied to the record. First, verify that the locality is correct by comparing the data to the label. Make corrections as needed and georeference if possible. If no corrections are needed, do not georeference, move to “Closed” with a concise explanation in “Georeferencing Remarks.”

*Example:* The locality reads “at the intersection of 1<sup>st</sup> Street and 3<sup>rd</sup> Street” in a particular town, and you know that these streets do not intersect, and have not done so historically, in that town.

5. *Geopolitical Divisions Wrong* – If geopolitical divisions (country, state, county, etc.) and the locality description do not match, it is possible that the label information is incorrect and therefore should not be georeferenced. First, verify that the transcribed locality is correct by comparing to the label image. Make corrections as needed and georeference if possible. If there are no corrections to make, do not georeference.

*Example:* “5 mi S of Perkasio, Philadelphia County” In this case, the collecting locality (5 miles south of Perkasio) would be in Montgomery County, and the point of reference (Perkasio) would be in Bucks County, neither of which are in Philadelphia County. In this case, do not georeference, move to Closed and include a concise explanation in Georeference Remarks (e.g., “not georeferenced, county contradicts locality”).

*Note!* In some cases, the county may be referencing the nearest named place, rather than the locality in which the specimen was collected, in which case georeferencing may be possible.

*Example:* “4 mi N of Elizabethtown, Lancaster County” In this case, the collecting locality (4 miles north of Elizabethtown) would be in Dauphin County, but Lancaster county appears to be associated with the point of reference (Elizabethtown) rather than the collecting locality itself. In this case, accept the coordinates and move the record to Stage 2.

### 3.4. Manual Georeferencing

When GeoLocate does not find a locality, and the locality does not fit any of the characteristics that would make it unable to be georeferenced (see Section 3.3), it may need to be manually georeferenced. Refer to Wiczorek et al. (2012) Georeferencing Quick Reference Guide, for a categorization of the different types of localities that can be manually georeferenced and instructions for how to do so for each. This guide can be found as a pdf on the MAM website or online (see References). Manual georeferencing can be done in various tools, including GeoLocate within Symbiota, the batch georeferencing tool in Symbiota, Google Maps, and Google Earth (see section 5.3 for more details on the last two). However, if you have repeat localities (i.e., the same locality for more than one record) that are to be manually georeferenced, it would be useful to enter the manual georeferencing results, no matter how you procure them, through the Georeferencing Batch Tool in the MAM portal. This will save you from having to enter the coordinates you determined for each record with the same locality individually.

### 3.5. CoGe: A Collaborative Georeferencing Tool

Developed as a collaborative interface for the GeoLocate platform, CoGe is a multi-user georeferencing software that allows one to set up georeferencing projects, manage users, track progress and review results.

Reasons why CoGe may be a useful tool for georeferencing your collection may include:

- similar/identical localities are automatically grouped by GeoLocate
- provides a way to pool records to be georeferenced *across* collections (for example, you could do all the MD records for multiple herbaria in MAM, while another MAM collaborator does all the NJ records for the same herbaria)

- administrators can invite anyone with a CoGe login to georeference in their collection (for example, this may be a useful way to have researchers planning to georeference your sheets anyway to get that data into your collection as part of the georeferencing process)
- administrators can assign particular records to a georeferencer by geographic area (e.g., county) or by taxon (e.g., *Acer rubrum*)
- georeferencers can skip records and leave a note as to why they skipped them
- administrators can review work before bringing it back into the Symbiota portal; corrections can be made and tracked in CoGe

Conveniently, this tool is available through Symbiota, obviating the need to move data through csv files, but currently, one must go through the SERNEC door to get to it. To get started, here are instructions to get to CoGe through Symbiota and move data into it:

1. Go to the SERNEC portal ([sernecportal.org](http://sernecportal.org)).
2. Log in with your MAM username and password.
3. Navigate to your collection as you would in the MAM portal.
4. Find and click on the Processing Toolbox under Administration Control Panel.
5. Click on the GeoLocate CoGe tab.
6. Halfway down the page, click on Login to CoGe. Create a username and password in CoGe via the Sign Up button. Once logged into CoGe, create a community (= georeferencing project) to deposit data into. Click on Dashboard at top of main page, then click on Create tab. Name the community, choose georeferencing options and select a privacy setting.
7. Return to SERNEC portal and click on Check Status, or refresh page, to get your community to show under Available Communities. Using the filters at the top of the page, choose which data<sup>4</sup> from your collection will go into CoGe for collaborative georeferencing (see Match Count for results of filtering). Select the Community you want to move data into and enter a name for the data source identifier, then click the Push Data to GeoLocate CoGe button to move it.

For further information and instructions, see the website (<http://coge.geo-locate.org/>) and the iDigBio resource page (<https://www.idigbio.org/content/coge-collaborative-georeferencing-demo-webinar>), which includes links to instructional videos.

#### 4. Review & Clean Data

Records that are georeferenced using the guidelines in this document should have “MAM Georeferencing Guidelines (2019)” entered in the field Georeference Protocol. This can be done by batch editing as follows. To find the relevant records to batch edit, perform a search with the following parameters in Edit Existing Occurrence Records:

Custom Field 1: Decimal Latitude IS NOT NULL

Processing Status: Stage 2

This step can wait until georeferencing is completely finished or can be repeated as necessary every few months or so. We typically include this designation in Georeference Protocol even for records that had coordinates recorded at time of collection because of the evaluation procedures included in the guidelines for whether or not to accept those coordinates.

---

<sup>4</sup> Note that only records without latitude/longitude can be moved out of Symbiota and into CoGe at present. To move already georeferenced records, one must use the transfer data by csv file format.

Unfortunately, at this time, there is no easy way to review georeferenced data to find errors. At a minimum we recommend reviewing that each georeferenced point is in the correct county by comparison to the County field. The basic idea is to download records by county from your collection (or multiple collections) and view them in a map software program that can show a county map layer. Here are instructions for how to do this record-by-county review in the program Google Earth.

1. If you do not already have Google Earth on your computer, download and install the software.
2. In the MAM portal, navigate to the herbarium of interest<sup>5</sup> scroll to the bottom of the page to the Extra Statistics box and click on Show Geographic Distribution. Then click on United States and your state of choice to get a county list. Click on the number in parentheses (= the number of records) after a county name to list the specimens for that county.<sup>6</sup>
3. Click on the tab at the top called Maps. Then click on the Create KML button. Save the resulting file.
4. Open Google Earth. In the Layers panel to the right, check at least the “2<sup>nd</sup> Level Admin Regions (Counties)” box to show county boundaries on the map.
5. Click File > Open, then select your downloaded kml file, then click Open. (If you prefer a straight-down view of your map, rather than a tilted view, press the R key once you have a map loaded.)

Now you should be able to see all your points from the county for which you downloaded data (with each point labeled with Collector + Collector Number) and the county boundaries together on one map. Review the localities to see that they are in the correct county and make a note of those that are not. There is limited information about the specimen available if you click on the map icon for that specimen. To see more information, click on the RecordURL in the specimen info box. This pulls up the complete specimen record within Google Earth for viewing. (To actually work with the record, you would need to copy the specimen url or the catalog number and pull it up in the MAM portal.) For any points you find that occur outside the county boundaries, follow up in the MAM portal to find the problem (remember to check that the original county/locality were transcribed correctly), and re-georeference, if needed.

If you are familiar with GIS, you can easily do this same procedure in ArcGIS (licensed software) or QGIS (free version). There is more power in using GIS than Google Earth because more data exists to facilitate your georeferencing review. For instance, you could use this same procedure to check that georeferenced points are within the correct boundaries for any other feature (e.g., preserve, state game lands) or named place (e.g., townships) for which there are boundaries available for GIS (e.g., shapefiles). [To retrieve data from your collection for particular places, search on key phrases (e.g., Nockamixon State Park) in Locality following the instructions in footnote 4.]

There is a “spatial module” for Symbiota that was developed through the macroalgae TCN, and is viewable on their portal (macroalgae.org), that would allow for the review of records by county (and, incidentally, records through time) in Symbiota online. Regrettably, this module has not been implemented in SEINet, and therefore the MAM portal, as of yet, though it may be in future.

---

<sup>5</sup> To download data for more than one herbarium at once: Click on the Search tab (near right side of tabs across top) in the MAM portal and choose Search Collections. Uncheck Deselect All then check off the collections from which you want to download data then click Search. Enter a county into the County box under Locality Criteria, then click List Display. Return to Step 3 in instructions to continue.

<sup>6</sup> If you save each kml with a standard name (e.g., “State\_County”), you could download all the records by county in your collection, then upload them all into Google Earth at once. Check and uncheck the appropriate files in the right-hand sidebar in Google Earth to toggle specific layers on/off as desired, or to see one county at a time.

## 5. Citations & Resources

### 5.1. Citations

Rios, N. 2018. GeoLocate. A platform for Georeferencing Natural History Collections Data.  
<http://www.geo-locate.org/>

Skema, C. and Anne Barber. Mid-Atlantic Megalopolis Digitization Standards and Training Manual, Version 2. <https://www.mamdigitization.org/reporting>

Symbiota Working Group. 2018. <http://symbiota.org/docs/symbiota-introduction/symbiota-overview/>

### 5.2. Resources about the Practice of Georeferencing

BioGeomancer Consortium. 2006. Guide to Best Practices for Georeferencing.  
<http://herpnet.org/herpnet/documents/biogeomancerguide.pdf>

iDigBio Georeferencing Resources. Various blogs, videos, webinars, etc...  
<https://www.idigbio.org/tags/georeferencing>

Wieczorek, J., D. Bloom, H. Constable, J. Fang, M. Koo, C. Spencer and K. Yamamoto. 2012. Georeferencing Quick Reference Guide.  
<http://www.herpnet.org/herpnet/documents/GeoreferencingQuickGuide.pdf>

Wieczorek, J. 2001-2015. MaNIS/HerpNet/ORNIS Georeferencing Guidelines.  
<http://manisnet.org/GeorefGuide.html>

### 5.3. Resources for Use in Georeferencing

**Table 3.** Resources for use in georeferencing.

Resource	Application	Access
All Trails	Locating coordinates for trailheads	<a href="https://www.alltrails.com/">https://www.alltrails.com/</a>
CoGe	Collaborative georeferencing tool	<a href="http://coge.geo-locate.org/">http://coge.geo-locate.org/</a> See Section 3.5 for more details.
GeoLocate	Primary application for georeferencing	<a href="http://www.geo-locate.org/">http://www.geo-locate.org/</a> Use in MAM portal, see Sections 3.1-3.2 for more details.
Google Earth*	Measuring distances	<a href="https://www.google.com/earth/download/gep/agree.html">https://www.google.com/earth/download/gep/agree.html</a>
Google Maps*	Locating coordinates and measuring distances	<a href="https://www.google.com/maps">https://www.google.com/maps</a>
USGS	Locating coordinates for named places	<a href="https://geonames.usgs.gov/">https://geonames.usgs.gov/</a>

\*Notes on use of Google Earth and Google Maps:

*Google Earth* - Download and install the application, then use these configurations.

First time set up: If you prefer a straight-down view of your map, rather than a tilted view, press the R key once you have a map loaded. Under Tools > Options and select Decimal Degrees under Show Lat/Long and Meters, Kilometers under Units of Measurement, then click OK. Then, in the left panel under Layers, check the box next to Borders and Labels. Uncheck all other layers. You may turn them back on depending on your georeferencing needs, but turning them off cleans up the map area,

making it easier to see what you are doing. Start the ruler tool by clicking on the ruler icon in the top navigation panel or by going to Tools > Ruler.

*Google Maps* – To get coordinates for a locality in Google Maps, right click on the locality pin and select “What’s here.” To measure in Google Maps, right click on the locality pin and select “Measure distance.”