This workflow was developed at an iDigBio workshop in January 2015. The most recent version is available at https://github.com/iDigBioWorkflows/FlatSheetsDigitizationWorkflows and https://www.idigbio.org/content/workflow-modules-and-task-lists.

Appendix S13. Module 13: Georeferencing

While georeferencing can be incorporated into the specimen digitization process at multiple points, this protocol assumes that the specimen is being georeferenced after complete transcription and digitization of the label data into a format from which a table can be generated. This will permit batch or collaborative georeferencing and reduce the possibility of separately georeferencing multiple specimens from the same locality.

Georeferencing can occur at the level of the individual specimen or with grouped specimens. Obviously, efficiency is greater when specimens from the same locality are grouped before or during the georeferencing process. Certain programs (e.g., GEOLocate, http://www.museum.tulane.edu/geolocate/; Symbiota, http://symbiota.org/) provide the means for grouping specimens for batch georeferencing.

This workflow is for georeferencing *legacy* locality data and assumes GEOLocate will be used to georeference specimens in batch. For information about recording locality information for specimens as they are collected in the field, see Module 14: Proactive Digitization. For information about Symbiota's batch georeferencing tool, see this webpage: http://symbiota.org/docs/batch-georeferencing-tool-2/.

Before embarking on the task of georeferencing, digitizing staff should review the various training materials that are available or consider participating in a training activity (See: http://georeferencing.org/training.html; https://www.idigbio.org/wiki/index.php/Georeferencing).

Task ID	Task Description	Explanations and Comments	Resources
T1	Separate specimens that have coordinates from those that do not.	For specimens lacking coordinates, batch or collaborative georeferencing should be performed. For specimen records that already include geographic coordinates, it is necessary to input the datum used, uncertainty measurement, and units into the digitized specimen information. If these data are not available it is recommended to check the verbatim coordinates via re-georeference in the descriptive locality starting at T4.	Specimen occurrence records with locality data.

T2	Clean and transform data for georeferencing.	The batch or collaborative functions of GEOLocate require a very particular CSV format (refer to the link). Sort the data appropriately (e.g., county, collector, year, collection number, etc.) so that specimens with identical localities are not repeatedly georeferenced independently.	Spreadsheet application. See: GEOLocate tutorial, http://www.museum.tulane.edu/geolocate/standalone/tutorial.html.
T3	Upload dataset into GEOLocate.	Batch or collaborative features of GEOLocate are most efficient for large datasets. These functions allow researchers to rapidly work through the separate occurrences, edit the uncertainty, and measure distances without leaving the GEOLocate interface. For batch GEOLocate, be sure to record the retrieval code at the bottom of the screen in order to access your data in the future.	GEOLocate.
T4a	Identify and verify locality in GEOLocate.	GEOLocate will estimate the geographic coordinates and precision of the specimen based on the written locality description, county, and state. The many features of GEOLocate (map layers, measurement tools, zoom features, etc.) should be explored to find the most appropriate tools for the locality information being georeferenced. For example, the Hybrid map layer is especially useful for identifying U.S. Forest Service road numbers. At any point in the georeferencing process, it may become obvious that a specimen cannot be georeferenced. In such cases, designate the occurrence as such in the georeferencing remarks field, describing the reason. Recommended is to develop a collection or institution-specific georeferencing policy that clarifies the conditions under which georeferencing for a particular specimen is to be undertaken. The policy should spell out the minimum amount of locality data required before georeferencing is attempted.	See: Wieczorek et al. (2012). https://www.idigbio.org/sites/default/files/workshop-presentations/geotrain/GeoreferencingQuickReferenceGuide20121008.pdf.

		Usefulness of georeferences at various resolutions (county, state, country) is the critical factor. See the Georeferencing Quick Reference Guide for information about how to select a coordinate location (i.e., place the marker) for various kinds of locality types.	
T4b	Identify locality with alternative resource.	Incorporation of resources beyond GEOLocate is also recommended to confirm or contest the estimated coordinate provided by GEOLocate. Alternative resources include Google Maps, Google Earth, Falling Rain, Getty Thesaurus, and NOAA interactive catalog, each of which has its benefits in identifying certain locality features. These tools should be used alongside GEOLocate to identify troublesome (historical, non-United States, water, etc.) localities and refined coordinates inputted into GEOLocate.	See: Georeferencing.org gazetteer resources: http://georeferencing.org/gazetteers.html. See: Wieczorek et al. (2012). https://www.idigbio.org/sites/default/files/workshop-presentations/geotrain/GeoreferencingQuickReferenceGuide20121008.pdf.
T4c	Evaluate spatial extent and uncertainty of the locality.	The spatial extent and uncertainty of the locality should be determined based on Biogeomancer Best Practices (see document link). The home institution should decide its own policy and protocol regarding preferential uncertainty type (point-radius, polygon, etc.) before georeferencing.	See: Chapman and Wieczorek (2006). http://www.gbif.org/resources/2809. See: Wieczorek et al. (2012). https://www.idigbio.org/sites/default/files/workshop-presentations/geotrain/GeoreferencingQuickReferenceGuide20121008.pdf.
T4d	Edit status of specimen.	If the specimen is confidently georeferenced with accompanying spatial extent and uncertainty, then the “Correct” button in GEOLocate should be selected, thus changing the “correction status” of the specimen from “no” to “yes.”	
T4e	Add georeferencing	The georeferencing protocol and/or georeferencing remarks fields should	

	protocol and georeferencing remarks.	be added to each specimen in GEOLocate regarding how the georeference was made (i.e., identified in Google Maps; latitude/longitude from specimen label, if coordinates are provided), how the uncertainty was determined, and any caveats about accuracy.	
T5	Export georeferenced data.	The georeferenced dataset can be exported from GEOLocate via the "File Management" link at the bottom of the screen.	
T6	Quality control visualization.	Check your georeferenced localities by visualizing the coordinates to ensure they are located in a reasonable locale.	GPS Visualizer (www.gpsvisualizer.com/). Google Earth.
T7	Check for completeness and quality.	As many Darwin Core georeferencing fields should be completed as possible and entered into the database. Example: georeferencedBy georeferencedDate georeferenceProtocol georeferenceSources georeferenceVerificationStatus georeferenceRemarks verbatimCoordinates verbatimLatitude verbatimLongitude verbatimCoordinateSystem decimalLatitude decimalLongitude geodeticDatum coordinateUncertaintyInMeters coordinatePrecision locality verbatimLocality	See: Darwin Core, http://tdwg.github.io/dwc/terms/index.htm#Location and http://tdwg.github.io/dc/terms/index.htm#GeologicalContext.

Literature Cited

Chapman, A. D., and J. Wieczorek (eds). 2006. Guide to Best Practices for Georeferencing. Global Biodiversity Information Facility, Copenhagen, Denmark. http://www.gbif.org/orc/?doc_id=1288.

Wieczorek, J., D. Bloom, H. Constable, J. Fang, M. Koo, C. Spencer, and K. Yamamoto. 2012. Georeferencing quick reference guide. Version: 2012-10-08. https://www.idigbio.org/sites/default/files/workshop-presentations/geotrain/GeoreferencingQuickReferenceGuide20121008.pdf.