

RESEARCH PAPER

Mapping Life – Quality Assessment of Novice vs. Expert Georeferencers

Elizabeth R. Ellwood*, Henry L. Bart, Jr.[†], Michael H. Doosey[†], Dean K. Jue[‡], Justin G. Mann[†], Gil Nelson^{*§}, Nelson Rios[†] and Austin R. Mast*

The majority of the world's billions of biodiversity specimens are tucked away in museum cabinets with only minimal, if any, digital records of the information they contain. Global efforts to digitize specimens are underway, yet the scale of the task is daunting. Fortunately, many activities associated with digitization do not require extensive training and could benefit from the involvement of citizen science participants. However, the quality of the data generated in this way is not well understood. With two experiments presented here, we examine the efficacy of citizen science participants in georeferencing specimen collection localities. In the absence of an online citizen science georeferencing platform and community, students served as a proxy for the larger citizen science population. At Tulane University and Florida State University, undergraduate students and experts used the GEOLocate platform to georeference fish and plant specimen localities, respectively. Our results provide a first-approximation of what can be expected from citizen science participants with minimal georeferencing training as a benchmark for future innovations. After outliers were removed, the range between student and expert georeferenced points was <1.0 to ca. 40.0 km for both the fish and the plant experiments, with an overall mean of 8.3 km and 4.4 km, respectively. Engaging students in the process improved results beyond GEOLocate's algorithm alone. Calculation of a median point from replicate points improved results further, as did recognition of good georeferencers (e.g., creation of median points contributed by the best 50% of contributors). We provide recommendations for improving accuracy further. We call for the creation of an online citizen science georeferencing platform.

Keywords: Benchmarking; biodiversity specimens; georeferencing; natural history collections; quality assessment; volunteered geographic information

Over the past decade the abundance of location-aware mobile devices has simplified recording of high-precision, high-accuracy geospatial data for the distribution of organisms. Several mobile apps are now available for this purpose (e.g., iNaturalist; iSpot; ebird); these contribute to the quality of citizen science databases (Spyratos and Lutz 2014). However, most biodiversity specimens collected prior to the 1990s do not have a latitude and longitude associated with them (Beaman and Conn 2003). This means that many of the world's three billion biodiversity specimens (Beach et al. 2010), including insects on pins, plants on sheets, and fish in jars—some collected as long as three centuries ago—

are not easily mapped. Therefore, their value as an historical baseline for research, education, and policymaking is limited (Cook et al. 2014; Hanken 2013).

Citizen science participants are playing an increasingly important role in transcribing specimen label data (Ellwood et al. 2015), but the expansion of georeferencing of specimen collection localities by public participants lags, partly owing to the dearth of online tools enabling georeferencing and the lack of experiments assessing the quality of the data produced. Here we present two experiments in which locality descriptions were georeferenced (assigned a latitude and longitude coordinate) by both expert and novice participants. We compare the data generated by the two groups and suggest downstream analyses to produce the most accurate locality estimates.

Georeferencing of historical localities is just one of many applications within the field of historical GIS (Gregory and Ell 2007). While we focus here on members of the public georeferencing biodiversity specimens, research in the digital humanities also has made important contributions to current georeferencing methodologies and technologies. For example Georeferencer, an online application

* Department of Biological Science, Florida State University, Tallahassee, FL 32306, United States

[†] Tulane University Biodiversity Research Institute, Belle Chasse, LA, 70037, United States

[‡] Florida Resources and Environmental Analysis Center, Florida State University, Tallahassee, FL 32306, United States

[§] Institute for Digital Information, Florida State University, Tallahassee, FL 32306, United States

Corresponding author: Austin R. Mast (amast@bio.fsu.edu)

designed to enable crowd-sourced rectifying of digital images of historic maps, has been modified and successfully implemented by numerous European institutions (Fleet et al. 2012). These efforts have resulted in tens of thousands of maps available online for increased discoverability, integration with modern map layers, improved visualizations, and a host of specialized research projects (Fleet et al. 2012; Holdsworth 2003; www.bl.uk/maps/georef/about.html). Like other fields, the digital humanities have turned to volunteers and crowd-sourcing to improve the rate at which historic documents are georeferenced (Offen 2012).

Volunteered Geographic Information (VGI) is a term coined in 2007 (Goodchild 2007) to recognize the fact that Internet-based media were incorporating geographic information wherever possible, including websites and mobile device apps for shopping, mapping, social connections, and weather (Sui and Goodchild 2011). VGI has grown tremendously over the last decade as evidenced by the millions of registered users on OpenStreetMap (openstreetmap.org; Haklay and Weber 2008)—a world map created and maintained by volunteers—and WikiMapia (wikimapia.org), a highly annotated world map with embedded links to related Wikipedia articles. OpenStreetMap also has a humanitarian arm of volunteers who are applying their geographical skills in poorly mapped parts of the world which are in need of aid, e.g., after the years-long rebellion in the Central African Republic and after the 2015 earthquake in Nepal (hot.openstreetmap.org).

Geotagging also has grown in popularity as text messaging systems, social media outlets, and photo sharing sites (in particular Flickr.com) have enabled users to include geographic information with these various media (Barve 2014; Kumar and Seitz 2014). Participation in, and demand for, this functionality illustrates a general public interest in working with geographic interfaces, expanding geographic data and improving freely available geographic information. Specific applications of geotagging have allowed researchers to track epidemic outbreaks (Lamos and Cristianini 2010), leverage the public's interest in visiting clean water bodies for improved water quality (Keeler et al. 2015), and improve epidemiology research (Doherty et al. 2011).

While research applications of VGI are relatively common (Sui et al. 2013), working with volunteers to add geographic information based on a textual description is relatively uncommon. In one of the few existing examples, volunteers added geographical information to social media posts to provide targeted and specific help to victims of the 2010 earthquake in Haiti (Meier 2012). Immediately after the earthquake, Haitian and college student volunteers in Boston, Massachusetts, scoured the web for social media posts related to the event and created a live map of the locations from where they were sent. Some of these posts had geographic information embedded in them, while others were textual descriptions of a location (i.e., “trapped under house at corner of Main and 1st”; Camponovo and Friendschuh 2014; Meier 2012) that needed to be given a latitude and longitude. Volunteers classified the posts

based on the type of aid that was needed and added them to the map; relief organizations then were able to use the live map to provide timely, appropriate help to individuals around the country.

Though less immediately urgent, the approach needed when georeferencing biodiversity specimens is similar to the above example. That is, citizen science participants read locality information in the form of short textual descriptions and transform that information into a latitude and longitude (i.e., a point on a map) and some measure of uncertainty, such as the radius of a circle. Biodiversity research specimens include a description of the locality that references political units (e.g., country, state, county); proximity to the nearest town or other geographical features; and/or the habitat (e.g., roadside, forest, lakeshore). Most descriptions require some interpretation and inference on the part of the georeferencer. The biodiversity research community previously established best practices for this type of work (Chapman et al. 2006), however, these practices were described prior to the recent expansion of VGI (Elwood et al. 2011; Goodchild 2007).

Georeferenced biodiversity specimens are crucial for many research applications including conservation (e.g., Miller et al. 2012; Rivers et al. 2011), estimating species ranges and extinctions (e.g., Boakes et al. 2010; Gotelli et al. 2012; Tingley and Beissinger 2009), habitat modeling (e.g., Fernández et al. 2015; Hope et al. 2013; Zhang et al. 2012), and natural resources management (e.g., Taylor et al. 2013). However, the level of accuracy and precision of georeferenced data impacts the quality of the downstream research (Graham et al. 2008; Rowe 2005). Taking advantage of the irreplaceable historical data provided by georeferenced biodiversity specimens will require a tremendous effort to georeference specimens currently in collections (Beach et al. 2010) using efficient methods leading to precise results (e.g., Guo et al. 2008).

Consider an example locality descriptions from the label of a plant specimen collected in 1927 in Highlands County, Florida, which reads “High pine land; Lake Stearns, Fla.” (**Fig. 1**). Turning this locality into a point on a map requires that a georeferencer find the town of Lake Stearns, determine where high pine habitat is likely to occur, and designate a point with a radius of uncertainty that encompasses the most likely collection location(s) of this specimen. To further complicate this process, habitat types and town names change over time. Since the time this specimen was collected nearly 90 years ago the town of Lake Stearns has changed its name to Lake Placid, and the high pine habitat where this specimen was collected may have ceased to exist. Even an expert georeferencer may have trouble as map layers usually reflect only current information, and finding historical town names and habitat types can be challenging. Also, specimen collection localities may be intentionally imprecise if a species is rare (e.g., to reduce illegal harvesting), and during some time periods and at some locations in the last three centuries, collectors were uncertain about precise locations because fine-scale maps and distinguishing features of the landscape were unavailable. Although many collection locality descriptions may be more straightforward than the one

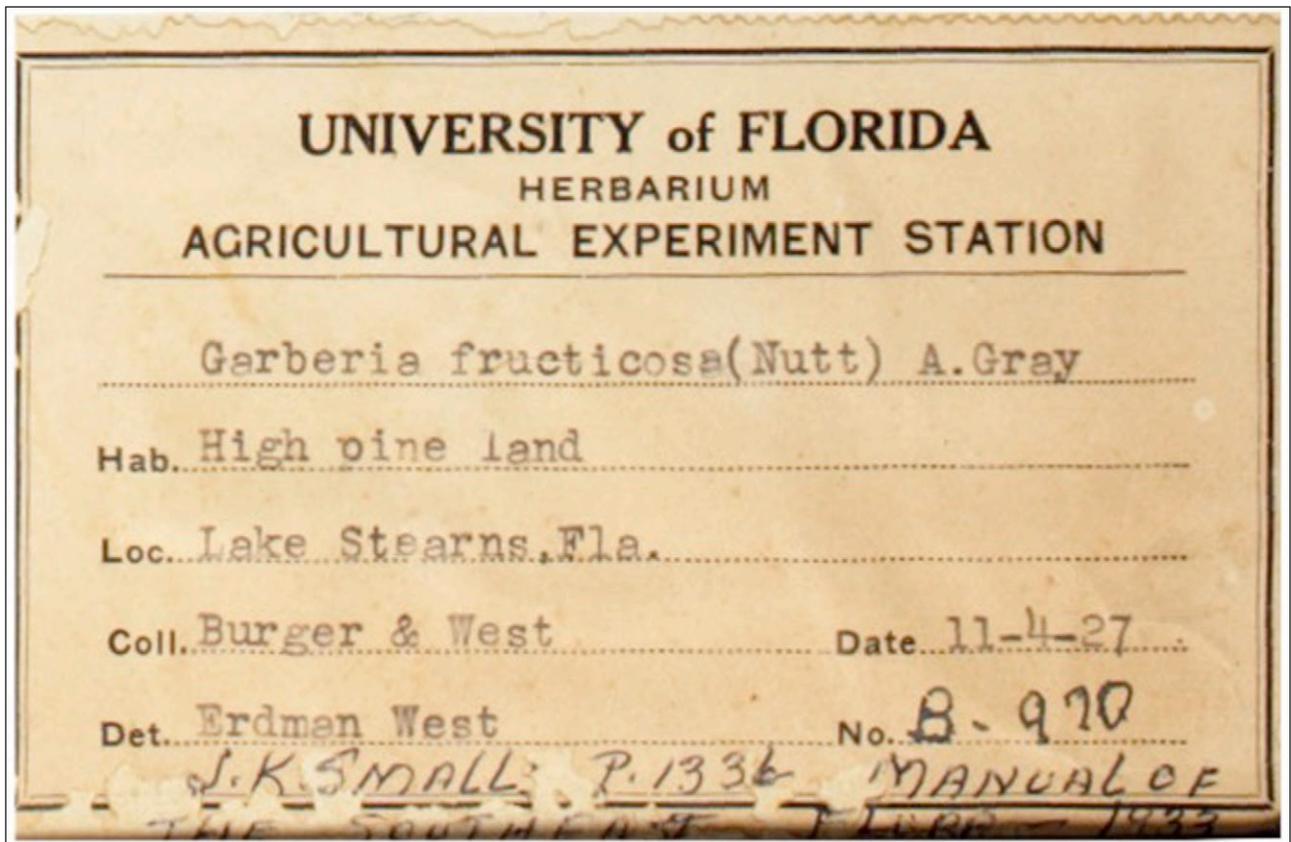


Figure 1: Label from a plant specimen from the Robert K. Godfrey Herbarium, Florida State University, Tallahassee, FL, US, demonstrating the potential challenges of georeferencing collection localities. In this case, the town has changed names since 1927, the locality description is imprecise, and the habitat is likely now residential development. Labels with such characteristics may be especially difficult for citizen science participants to georeference without local knowledge.

provided in this example, considering the breadth of heterogeneity in locality descriptions, can citizen science participants contribute accurate and appropriately precise specimen georeferences?

To investigate this question, we engaged undergraduate students as a proxy for the general population of citizen science participants. While we do not have data demonstrating that these students are comparable to the general citizen science community, they are a subset of the general population and represent a range of abilities, levels of innate interest, and prior experience with geographical information and biodiversity research. We chose to use students so that we could generate sufficient data in the absence of an established citizen science georeferencing platform and community. We asked:

1. How accurate are student georeferencers compared to automated georeferencing software and experts? Does student involvement improve on the accuracy of a georeferencing algorithm?
2. What method is most effective at estimating an accurate consensus georeference from replicate points for the same collection locality? Is the consensus generated in this way more accurate than the individual points?
3. How do the best georeferencers compare to the group as a whole? That is, is it useful to only

consider the points produced by the most accurate georeferencers?

Methods

To address our research questions we conducted two experiments in which undergraduate students and experts georeferenced the same collection localities. The two experiments differed in the spatial distribution of collection localities (seven states in the USA vs. Florida's Apalachicola National Forest), the biology of the organisms (fish vs. plants), and the number of student georeferences for each locality (1–2 vs. 6–15 respectively). We addressed question 1 with both datasets and questions 2 and 3 with the many-georeferences-per-location dataset.

Each of the experiments relied on GEOLocate software (www.museum.tulane.edu/geolocate/), which uses an automated georeferencing algorithm to make the human georeferencing more efficient. The algorithm interprets strings of text and provides a suggested point location and radius of uncertainty. GEOLocate displays the most likely point as a green dot and shows red dots for other possible, though less likely, points based on the GEOLocate algorithm. A user can choose one of these suggestions or create another point. GEOLocate also includes features that allow a user to view different map layers, expand the screen, zoom and pan, mark a spot, measure, and save a

point. All participants used GEOLocate to assess, navigate, and extract spatial information.

Fish experiment: Thousands of fish localities each georeferenced by one or two students

In the first experiment, 3,372 U.S. fish collection localities from Fishnet2 (fishnet2.net/aboutFishNet.html) were each georeferenced by one (or occasionally two) undergraduate student georeferencers at Tulane University (New Orleans, Louisiana, USA) using GEOLocate's Collaborative Georeferencing platform (museum.tulane.edu/geolocate/community; CoGe). The data were grouped into seven state datasets and distributed among 11 students (undergraduate students in Natural Resource Conservation and Biodiversity Informatics classes taught at Tulane) and eight trained and experienced project technicians, such that each dataset was georeferenced by at least one student and at least one trained, experienced technician. Students and technicians corrected the geolocation recommended by GEOLocate when necessary and saved the latitude and longitude of that chosen location. Student training involved a 50-minute overview on georeferencing biodiversity data followed by demonstrations on using GEOLocate and CoGe. The technicians were hired specifically to georeference fish specimen localities as part of a research grant. They received two days of training, encompassing basic geographic principles, georeferencing methodologies and standards, and project protocols. Many of them had GIS experience prior to the project, and all of them had months of experience georeferencing localities in the project by the time of the experiment.

At Tulane, data processing and analyses were conducted using PostgreSQL 9.3, PostGIS 2.1, Microsoft Access 2010, Microsoft Excel 2010, and Microsoft Excel 2013. Distances between student and expert points and distances between most highly suggested point in GEOLocate and expert points were compared. Records that were not resolvable by GEOLocate were excluded from GEOLocate comparisons. Because we had only one or two student results for each technician result for each locality in the fish dataset, we could not compute means and medians across student results as in the plant experiment.

Plant experiment: Hundreds of plant localities each georeferenced by many students

In the second experiment, 270 plant collection localities from Florida's Apalachicola National Forest (ANF) each were georeferenced by 6–15 students at Florida State University (FSU, Tallahassee, Florida, USA) using GEOLocate's standard online platform. The plant collection locality description were taken from the database of FSU's Robert K. Godfrey Herbarium (www.herbarium.bio.fsu.edu). Each student was provided an Excel worksheet with collection information parsed into columns: Specimen barcode, scientific name, country, state, county, and locality description. The locality description was an aggregation of entries in the following of the herbarium's database fields: Nearest Named Place, Special Geographic Unit, Verbatim Directions to Locality, and Habitat. An example is "Bristol, Apalachicola National Forest by Fla Rt. 12, S of Bristol,

Apalachicola National Forest, just within boundary, long-leaf pine savanna." An additional column contained links that took the student directly to the GEOLocate website with the specimen's locality description preloaded in the interface. The full Excel file had 17 different worksheets, each listing 16 specimens (with the exception of the last worksheet which had only 14 specimens).

Each of 154 Florida State University junior and senior undergraduate students enrolled in the course Plant Biology was assigned one worksheet (i.e., 16 or 14 specimen localities) from within the full file to georeference. As a class, students were provided with both a 30-minute training session and written instructions that included a step-by-step guide for augmenting the Excel file with a latitude and longitude (but not a measure of uncertainty) obtained from their work using GEOLocate. Although each worksheet was assigned to the same number of students, some students did not follow directions, so certain worksheets were completed more frequently than others. In the end, each specimen was georeferenced 6–15 times (mode = 8, median = 9).

When a student followed a specimen's link to GEOLocate, they were asked to use GEOLocate's automated georeferencing algorithm (a button "Georeference") to produce suggested points, then they could pan, zoom, and open other map layers to show different features, including political boundaries, streets, and aerial photos, until they found the closest approximation of the textual description. Then they cut and pasted the latitude and longitude into Excel. Completion of these tasks, regardless of accuracy, earned the student credit for the required assignment. However, students could opt out of the experiment by choosing not to complete an Institutional Review Board-approved waiver. Students were given one week to complete the assignment; during that time they could email one of us (GN) for guidance or help.

Independent of the student work, two local botanists with extensive collecting experience in ANF volunteered to also complete the georeferencing tasks. As local experts, they were familiar with habitat types in the ANF, specific plant populations, favored collection areas, and field collection protocols. This knowledge provided them the advantage over students of being able to more easily interpret and georeference label information. These individuals included a radius of uncertainty with their georeferences and made note of challenging or vague locality descriptions. The experts produced one point for each specimen, which henceforth are referred to as "expert" points.

A small subset of student points in the plant dataset were interpreted as outliers and were removed from the dataset. Such errors included latitude and/or longitude of 0, positive or negative latitude or longitude when the opposite was appropriate for the hemisphere, values that were incomplete, and values that were placed at the exact centroid of the nearby town of Apalachicola (representing an occasional mistake by the GEOLocate algorithm that students did not always correct; the town lies outside of the boundaries of ANF). We consider this data-cleaning step to be a reasonable approximation of what can be

done by any project doing georeferencing with citizen science participants, and are not using any special knowledge of the expert points at this step. Analyses were conducted with the remaining points in QGIS version 2.6.1 Brighton (QGIS Development Team 2014), Environmental Systems Research Institute's ArcGIS version 10.2 (Environmental Systems Research Institute 2014), and R statistical software version 3.1.1 (R Core Development Team 2014).

We calculated distance statistics between the expert point and points generated by students for each collection locality, including mean distance of student points and minimum and maximum distance of student points. For these plant experiment data, we calculated a mean and median georeferenced point for each collection locality from the replicate student points using ESRI's ArcMap spatial statistics tools Mean Center and Median Center, respectively. The Mean Center is simply the average X and average Y coordinate among all the points, while the Median Center tool utilizes an iterative algorithm to calculate the point that minimizes the Euclidian distance among all the student points for a given specimen record. The median point gives less weight to anomalous georeferences. For comparison, we also calculated the distance between the expert point and those suggested as most likely by the GEOLocate algorithm.

Individual students were evaluated for accuracy by comparing their mean distance from expert points (as measured using uncertainty radii for the specific specimens) for all specimens georeferenced by that individual. To determine the increased accuracy brought about by removing the least accurate georeferencers, we re-ran some of the analyses by first excluding 19 students whose complete set of georeferenced points averaged 100 uncertainty radii or greater from the expert's points, and then by excluding the bottom half (least accurate) of georeferencers. The first exclusion removes those participants who are perhaps least likely to contribute to a

citizen science project requiring this skill set, given their poor aptitude for it or their poor engagement in the activity. The second left us with a proxy for those members of the public who are devoted to a citizen science project and likely to become experienced in a way that becomes recognizable to the project. A disproportionate percentage of online tasks often are completed by a very small number of committed citizen science participants (Eveleigh et al. 2014).

Results

How accurate are student georeferencers?

Fish experiment—Eleven students produced 4,433 georeferences for 3,372 localities (1,061 localities georeferenced twice). The mean distance of student points from those of expert georeferencers ranged from 1.5–75.5 km (mean = 21.3 km). We defined outliers as student points that were greater than two standard deviations from the overall mean displacement of each student's result from the expert result; outlier distance ranged from 13–1884 km across all determinations. Georeferences with greater than a 25 km deviation were typically placed in the wrong county and/or state, and should be detectable through data validation routines involving spatial queries against administrative units in the absence of expert points. Numbers of outliers ranged from just 0–17 georeferences (mean = 6.5) per student. Excluding outliers, per-student mean distances between student and expert georeferencer determinations decreased to 0.9–40.7 km (overall mean = 8.3). Forty percent of student georeferences were within 0.5 km of the expert points, 53% were within 1 km, and 81% were within 5 km (**Fig. 2**). Considering the uncertainty radius assigned by the experts, 71% of student points were within one uncertainty radius of the expert, and 90% were within 10 (**Table 1**).

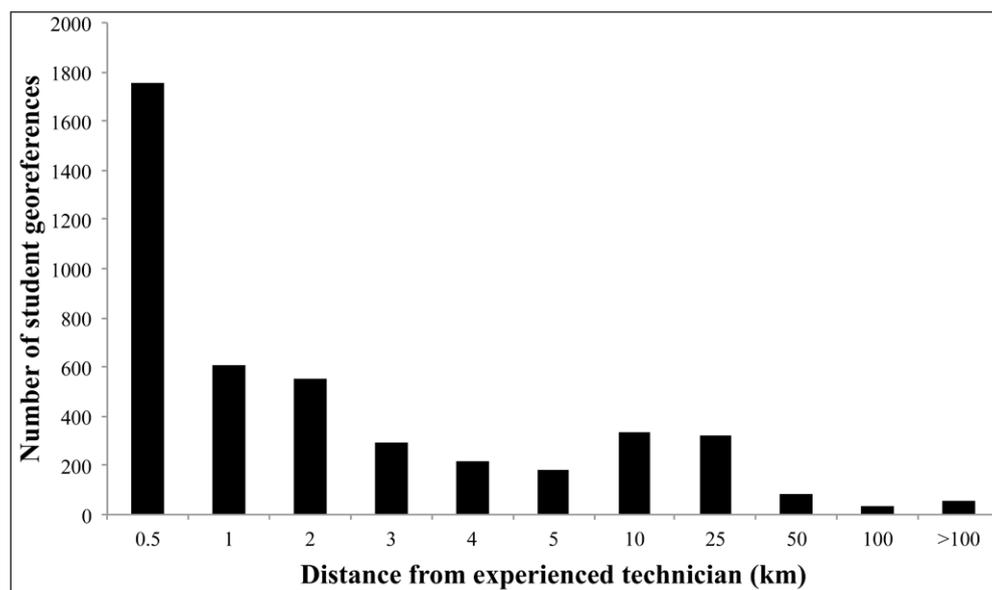


Figure 2: Distribution of the distance of student georeferences from expert points in the fish experiment at Tulane University with outliers removed.

We found that involving students in the process increased the percentage of points within each of the uncertainty radii cut-offs (**Table 1**; e.g., 71.07% vs. 49.09%, respectively, within 1 uncertainty radius as assigned by the expert georeferencers) and each of the absolute distance cut-offs less than the 10,000 meter cut-off (**Table 2**).

Plant experiment—A total of 2,425 georeferences were produced by students, and after removing outliers, 2,408 (99%) remained. The mean distance between student points and the expert point for each collection locality ranged from 0.18–37.08 km, with an overall mean student distance from the respective expert point of 4.62 km.

Points compared to expert points for each experiment	<1 UR	<2 URs	<5 URs	<10 URs	<100 URs
Individual Student Points—Fish (n = 4,433)	3260 (71.07%)	3546 (77.92%)	3897 (83.56%)	4037 (90.30%)	4367 (98.39%)
Individual Student Points—Plant (n = 2,408)	365 (15.16%)	627 (26.04%)	1070 (44.44%)	1428 (59.30%)	2254 (93.60%)
GEOLocate—Fish (n = 4347)	2134 (49.09%)	2590 (59.58%)	3062 (70.44%)	3463 (79.66%)	4152 (95.51%)
GEOLocate—Plant (n = 251)	32 (12.75%)	56 (22.31%)	101 (40.24%)	133 (52.99%)	223 (88.88%)
Mean for all Students—Plant (n = 270)	33 (12.22%)	56 (20.74%)	110 (37.04%)	155 (57.41%)	252 (93.33%)
Median for all Students—Plant (n = 270)	49 (18.15%)	79 (29.26%)	136 (50.37%)	190 (70.37%)	266 (98.52%)
Single Best Student Point for each Locality—Plant (n = 254)	99 (38.98%)	138 (54.33%)	204 (80.31%)	223 (87.79%)	254 (100.00%)
Median for Students Minus Worst 19—Plant Experiment 3 (n = 270)	63 (23.33%)	84 (31.11%)	139 (51.48%)	188 (69.63%)	269 (99.63%)
Median for Best Half of Students—Plant (n = 254)	52 (20.47%)	85 (33.46%)	141 (55.51%)	183 (72.05%)	254 (100.00%)

Table 1: Comparison of student points, consensus student points (using mean and median), and GEOLocate automated points to expert points measured by uncertainty radius (UR) for the fish and plant experiments. Because relatively few of the collection locations in the fish experiment were georeferenced by multiple students, we do not report comparisons with the consensus student points for that experiment.

Points compared to expert points for each experiment	Within 100 meters	Within 500 meters	Within 1000 meters	Within 5,000 meters	Within 10,000 meters	Within 20,000 meters
Individual Student Points—Fish (n = 4,433)	791 (18.34%)	1,314 (29.20%)	1,786 (40.47%)	2,941 (66.57%)	3,399 (76.92%)	3,782 (86.66%)
Individual Student Point—Plant (n = 2,408)	133 (5.52%)	464 (19.27%)	898 (37.29%)	1,698 (70.51%)	1,950 (80.98%)	2,182 (90.61%)
GEOLocate—Fish (n = 4347)	399 (9.17%)	937 (23.15%)	1439 (33.10%)	2810 (64.64%)	3372 (77.57%)	3858 (88.75%)
GEOLocate—Plant (n = 251)	19 (7.57%)	43 (17.13%)	96 (38.25%)	160 (63.75%)	190 (75.70%)	214 (85.26%)
Mean for all Student Points—Plant (n = 270)	50 (18.52%)	196 (72.59%)	222 (82.22%)	250 (92.59%)	266 (98.52%)	269 (99.63%)
Median for all Student Points—Plant (n = 270)	15 (5.56%)	69 (25.56%)	132 (48.89%)	221 (81.85%)	246 (91.11%)	267 (98.89%)
Single Best Student Point for each Locality—Plant (n = 254)	110 (43.30%)	163 (64.17%)	195 (76.77%)	237 (93.30%)	249 (98.03%)	253 (99.06%)
Median for Students minus Worst 19—Plant (n = 270)	64 (23.70%)	123 (45.55%)	163 (60.37%)	230 (85.19%)	253 (93.70%)	266 (98.52%)
Median for Best Half of Students—Plant (n=254)	60 (23.90%)	102 (40.64%)	150 (59.76%)	218 (86.85%)	240 (95.62%)	251 (98.82%)

Table 2: Comparison of student points, GEOLocate automated points, and median of student points to expert points measured by absolute distance for the fish and plant experiments. Because relatively few of the collection locations in the fish experiment were georeferenced by multiple students, we do not report comparisons with the consensus student points for that experiment.

To make the comparison between use of the automated georeferencing algorithm of GEOLocate alone and the additional involvement of the student georeferencers, we narrowed the number of collection localities to 251 because GEOLocate's suggested points for the other specimens were returned as errors. The most successful consensus georeferencing method (use of the median point for the replicate student points) places a greater proportion of points within the uncertainty radii thresholds than the GEOLocate-suggested point (**Table 1**). When measuring that distance in meters, the median point outperforms GEOLocate alone, except at a cut-off of 100 m (where GEOLocate alone has a slight advantage; **Table 2**).

Which method is most effective for producing an accurate consensus georeference?

For the plant data, use of the median georeferenced point as a consensus of replicate student georeferences is better than the mean georeferenced point at each of several uncertainty distances from the expert point (e.g., 12.22% of the mean points and 18.15% of the median points are within 1 uncertainty radius of their expert point; **Table 1**). Unless otherwise indicated, we will use the median georeferenced point as the standard for comparison of the consensus point with the expert point.

The same is true when we consider distance from the expert point using absolute distance (**Fig. 3**). For more than half of the student points in the plant experiment (58.60%; 1411 of 2408 points), the median point for a collection locality is at least 10 m closer to the expert point than the individual student point itself. About a quarter of the student points (25.83%; 622 points) are at least 10 m closer to the expert points than the median point

(**Table 2**). The remainder have similar distances to the expert point as the median point.

Is it useful to differentiate data based on georeferencer performance?

About 39% (99 of 254) of the single best student points for a collection locality are within one uncertainty radius of the expert point for that locality (**Table 1**), and about 43% of the single best student points are within 100 m of the expert point (**Table 2**). Examining the 99 single best points within one uncertainty radius we found that 48 (31%) of the 154 students contributed to them and just four students (3%) were responsible for 24 of those points.

We removed 19 of the 154 students contributing to the plant experiment using our threshold for identifying the least talented or motivated georeferencers, reducing the number of georeferenced points from 2408 to 2095 and the number of localities from 258 to 254. Using this reduced data set, the percentage of localities within one uncertainty radius of the expert increased from 18.15% with the full dataset to 23.33% (**Table 1**). Similarly, the percentage of localities that fell within 100 meters of the expert point increased from 5.56% with the full dataset to 23.70% with the reduced dataset (**Table 2**).

When we included only the best 74 (48%) of the plant georeferencers (1185 points), the distance of the median points calculated from the experts as measured by uncertainty radii was improved from the results of the full dataset, but not strikingly (e.g., 18.15% of the medians are within one uncertainty radius for the whole dataset vs. 20.47% for the subset; **Table 1**). Looking at improvement based on the absolute distance, however, shows a marked improvement (e.g., 5.56% of the medians are within

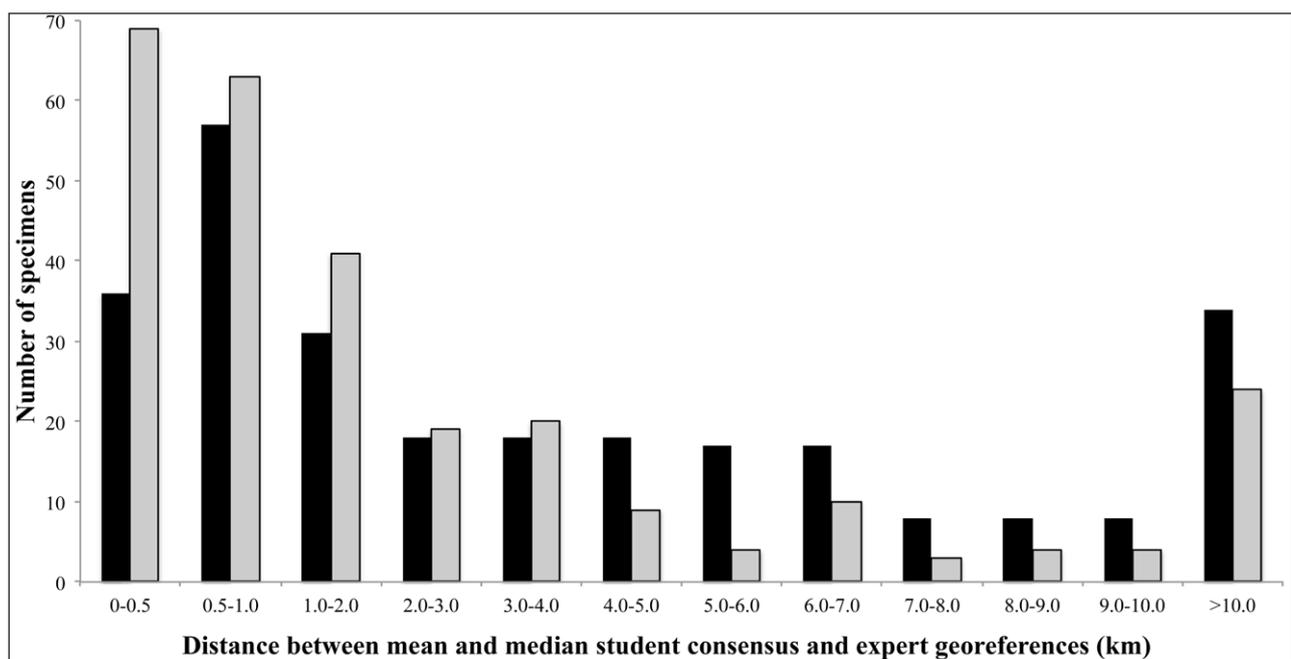


Figure 3: Distribution of the distance between mean (black bars) and median (gray bars) consensus of student replicate georeferences from the expert points in the plant experiment at Florida State University with outliers removed.

100 m for the total dataset vs. 23.90% of the medians for this subset vs.; **Table 2**).

Discussion

Our results provide a first approximation of what can be expected from citizen science participants with minimal georeferencing training. This is a valuable contribution, for while OpenStreetMap (Haklay and Weber 2008) and WikiMapia (wikimapia.org) have demonstrated enthusiasm for volunteered geographic information (Goodchild 2007), we are not aware of studies that have assessed the quality of citizen science georeferencing of collection localities for biodiversity specimens or, more generally, of points contributed by georeferencing novices using locality descriptions (e.g., as done by Meier 2012 in another domain). We consider the results encouraging and suggest that they might serve as a benchmark against which to compare future changes to the process, several of which we suggest here.

Our use of undergraduate students as proxies for the general citizen science population, in the absence of an established georeferencing citizen science platform and community, merits further discussion. Coleman et al. (2009) present a hierarchy of volunteer participation in the context of contributing geographic data. By their definitions, we expect our student volunteers to mostly be neophytes—“an individual without a formal background in a subject, but who possesses the interest, time, and willingness to offer an opinion” (page 338). Whether the potential population of citizen science participants who would contribute data in this way represents a similar fraction of neophytes remains unanswered by our study. Potentially a greater fraction of those who would be motivated to contribute, and possibly some of our more experienced undergraduate volunteers, would qualify as expert amateurs—“someone who may know a great deal about a subject, practices it passionately on occasion, but still does not rely on it for a living”—as would our expert volunteers in the plant experiment. (Our experts from the fish experiment would qualify as expert professionals in Coleman et al.’s scheme—“someone who has studied and practices a subject . . . [and] relies on that knowledge for a living.”) By Coleman et al.’s estimation, and further analysis by Lauriault and Mooney (2014), “expert amateurs” may be the most productive volunteer contributors of geographic information, although positive and negative motivations vary across projects and can influence relative involvement of a group. Targeting expert amateurs, or educating neophytes to become expert amateurs, in the biodiversity community might be an effective strategy for increasing contributions and improving their quality beyond that reported here. Expert amateurs might be found as members of native plant societies, entomological clubs, sportsmen’s groups, online communities such as iNaturalist (inaturalist.org), and conservation and environmental organizations. Members of historical societies may provide additional local knowledge and a familiarity with regional geographic and landscape features. Future research on the topic could benefit from including a broader demographic of citizen science participants

in experiments, along with additional methods such as surveys, to understand the advantages and limitations to working with each of these groups.

Despite large differences in the spatial extent of the areas considered in the experiments (seven states in the US vs. a national forest) and the biology of the organisms (fish in aquatic habitat vs. plants in, mostly, terrestrial habitat), the experiments produced strikingly similar average distances between student- and expert-contributed points (8.3 km with a range of 0.9–40.7 km and 4.6 km with a range of 0.2–37.1 km, respectively). However, when the distance is measured by uncertainty radii assigned for each collection locality by the experts, differences emerge. Relatively more of the contributed fish georeferences (71%) are within an uncertainty radius of the expert point than the plant georeferences (15%), perhaps because the extent of fish habitat is more easily identified on a map than that of plants and there is often relatively less of it. Also, the relatively larger uncertainty radii of the fish experiment (expert mean = 4,136 m, range = 0–457,118 m) than the plant experiment (mean = 1,054 m, range = 16–21,095 m) simplified the process for students to place a point within the uncertainty radius of the expert in that experiment.

Creation of a consensus point from replicates for a collection locality improved upon the overall percentage of points within one uncertainty radius in the plant experiment (the fish experiment did not consistently replicate) when the consensus was produced as the median point, but not the mean point (**Table 1**). The median is less sensitive to outliers and makes more sense than the mean for building consensus in this context. We do not address the relationship between number of replicates used to produce the median and the median’s accuracy here, but the relationship has clear importance when designing efficient citizen science projects in the domain. We expect a plateau above which more replicates do not improve accuracy of the median and therefore might represent wasted effort if other statistics are not also being estimated with the additional points. We expect that the location of such a plateau will vary from project to project for reasons discussed above (habitat requirements differ, as do typical sizes of uncertainty radii), and that location needs to be determined in a pilot study specific to that dataset until patterns begin to emerge across datasets. The additional points beyond those needed to improve the median might be important if used to estimate a measure of uncertainty for the locality if there is a relationship between the spread of points and the uncertainty that an expert might assign the locality (e.g., as an uncertainty radius or polygon; Chapman 2006). The relationship between spread and uncertainty might plateau at a different place than the accuracy of the median.

The accuracy of the data clearly improved beyond that produced using the automated GEOLocate algorithm when students were part of the workflow. The percentage of GEOLocate-generated points within an uncertainty radius of the expert points was improved upon by the students in both experiments (e.g., 12.75% vs. 15.16% within 1 uncertainty radius for the plants; **Table 1**), and

even more so when the median was calculated (18.15%). Note that the GEOLocate algorithm may have provided an important step in the student and expert contributions, especially in the fish experiment where the spatial extent of possible localities was very large. We actually cannot say whether the involvement of a georeferencing algorithm improved or reduced the accuracy of student points, because the experiment did not make that contrast. Future studies may wish to include an additional experiment that determines accuracy of citizen science participants in the absence of an algorithm. Further consideration of the topic, particularly by researchers in the field of human computation and machine learning, could investigate how the automated georeferencing algorithm could be improved by closing the loop—providing feedback to it in the form of citizen-science contributed data.

While the median-point consensus of replicates represented an improvement on the percentage of individual points within threshold numbers of uncertainty radii (e.g., 18.15% vs. 15.16% within 1 uncertainty radius for the plants; **Table 1**), the fact that the single best point for each locality is even more often within those thresholds (38.98% within 1 uncertainty radius for plants; **Table 1**) invites the question: are there ways to assess the likelihood that a contributed point is the best for a collection locality in the absence of expert points for all collection localities? One way that this might be accomplished is to assess the overall performance of georeferencers, assigning them reputation scores that reflect attributes such as success with localities for a handful of points that experts have georeferenced. A likelihood of success with such an approach is suggested by the fact that the 99 single best points within one uncertainty radius for plants were contributed by 31% of contributors (and not 65%, which would be one best per each of 99 of the 154 total students). Furthermore, a quarter of those 99 points were contributed by just four students.

We also looked at this relationship in another way, asking if the accuracy of the median point improves when data from only the best georeferencers are considered. In the case of thresholds of uncertainty radii, the percentages improved at most thresholds, but generally not dramatically (e.g., 50.37% at a threshold of 5 uncertainty radii for all georeferencers, 51.48% with exclusion of the 19 worst georeferencers, and 55.51% with the exclusion of the worst half of georeferencers; **Table 1**). The improvement is most striking, though, when the absolute distance of median from expert point is considered at low thresholds (e.g., 5.56% within 100 meters for all georeferencers and 23.70% and 23.90% with exclusion of 19 worst and worst half, respectively). This relationship can become especially relevant when the fitness for use depends on a precision within some absolute distance. For example, considering global latitudinal diversity gradients, modeling species distributions, and relocating a population are three activities that typically require increasingly precise data.

Hunter et al. (2013) provide a case study of an implementation involving data validation and trust metrics for improving the quality and measuring the reliability of citizen science data within Coral Watch (www.coralwatch.org).

A similar approach could be used to develop a weighted index of reputation based on some combination of (1) total number of user contributions, (2) frequency of user contributions, (3) geospatial deviation from known results, and (4) geospatial deviation for identical localities from users with higher reputation. Liu and Liu (2015) demonstrate a learning algorithm that can assess the quality of crowd-sourced data and provide results from only the strongest combination of contributors. The ability to sort “good” data from “bad” data, in an environment where the correct information is not known at the start, has obvious applications to the field of citizen science georeferencing, and we anticipate incorporating techniques similar to this in future work.

It is important to realize that, as illustrated in **Fig. 1**, there are specimens for which a precise georeference is not warranted and for which the actual collection locality is obscured by the changes of time. For example, 23% of the single best points for the plant localities were not within 1 km of the expert point, despite there being 6–15 replicates for each. Based on the plant dataset, types of labels that resulted in large discrepancies between expert and student points included these cases: a) Directional labels that do not specify how the distance is measured. For example, in the case of “Sumatra flatwoods pond, 16 miles N of Sumatra, flatwoods pond,” students measured 16 miles due north, while the experts followed the main road out of Sumatra, which veered to the northeast. This was a common problem, with three of the ten most poorly placed student points falling into this category; and b) Labels with overly general or contradictory information. For example, in the case of “4 miles NE of Sumatra, by Fla. Rt. 379,” there is likely an error because Route 379 runs in a northwesterly direction from Sumatra. The issue of flagging collection localities that are likely to fall into this category for georeferencing by experts or even the original collector (if still living) merits future consideration. Collection localities could perhaps be classified algorithmically with natural language processing into those requiring triage of this type to make more efficient citizen science engagement for georeferencing.

Finally, we recognize that potentially large improvements in accuracy could be gained with a dedicated citizen science platform for georeferencing of this kind. The georeferencing software packages that we used for the experiments were not created with neophyte contributors (*sensu* Coleman et al. 2009) in mind, and could be tailored to them to create more support for their activities (e.g., directions, feedback from simple data validation steps, a forum for discussion of issues) and that of the data curators, who could use reputation scores for data processing. Such a platform has been proposed for development (but not yet funded) as a special crowd-sourced georeferencing addition to GEOLocate’s suite of georeferencing software. The need for such a service is increasing as more and more pre-GIS locality records for the world’s billions of biodiversity specimens are digitized. The results of the present study suggest that novice georeferencers are capable of performing this task.

Acknowledgements

Our sincere thanks to the students and technicians who volunteered their time to georeference specimen labels, in particular FSU experts Patricia Stampe and Robin Kennedy. Thanks to Chris Buddenhagen for technical assistance and advice, and to FishNet PI's and technicians Diego Barroso, Lorenzo Benito, Andy Bentley, Melissa Casarez, Dave Catania, Paul DeSalles, Rick Feeney, Bill Fink, John Friel, Justin Grubich, Dean Hendrickson, Mariangeles Arce Hernandez, John Lundberg, Amy McCune, Theresa Lorraine McInnes, Douglas Nelson, Larry Page, Laura Porturas, Megan Roberson, Rob Robins, Luiz Rocha, Mark Sabaj, Leo Smith, Alexandra Snyder, Chris Taylor, Christine Thacker, Tom Turner, Michelle Vanderwel, Rachel Vinsel, Mark Westneat, Katy Wichman, and Ed Wiley. The Tulane experiment was supported by the National Science Foundation (Georeferencing U.S. Fish Collections: A Community-based Model to Georeferencing Natural History Collections (Collaborative), DBI 1202953).

The FSU experiment was supported by iDigBio, NSF's National Resource for Advancing Digitization of Biodiversity Collections. iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement no. EF-1115210). This material is based on work supported by the National Science Foundation under Cooperative Agreement no. EF-1029808. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Competing Interests

The authors declare that they have no competing interests.

References

- Barve, V., 2014. Discovering and developing primary biodiversity data from social networking sites: A novel approach. *Ecological Informatics*, 24: 194–199. DOI: <http://dx.doi.org/10.1016/j.ecoinf.2014.08.008>
- Beach, J., Blum, S., Donoghue, M., Ford, L., Guralnick, R., Mares, M., Thiers, B., Westneat, M., Wheeler, Q., Wiegmann, B. and the Network Integrated Biocollection Alliance, 2010. A strategic plan for establishing a network integrated biocollections alliance. Available at <http://dig-biocol.wordpress.com/brochure>.
- Beaman, R. and Conn, B., 2003. Automated geoparsing and georeferencing of Malesian collection locality data. *Telopea*, 10: 43–52. DOI: <http://dx.doi.org/10.7751/telopea20035604>
- Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K. and Mace, G.M., 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol*, 8: e1000385. DOI: <http://dx.doi.org/10.1371/journal.pbio.1000385>
- Camponovo, M.E. and Friendschuh, S.M., 2014. Assessing uncertainty in VGI for emergency response. *Cartography and Geographic Information Science*, 41: 440–455. DOI: <http://dx.doi.org/10.1080/15230406.2014.950332>
- Chapman, A.D., Wieczorek, J. and BioGeomancer Consortium, 2006. Guide to best practices for georeferencing. Global Biodiversity Information Facility.
- Coleman, D.J., Georgiadou, Y. and Labonte, J., 2009. Volunteered geographic information: The nature and motivation of producers. *International Journal of Spatial Data Infrastructures Research*, 4: 332–358.
- Cook, J.A., Edwards, S.V., Lacey, E.A., Guralnick, R.P., Soltis, P.S., Soltis, D.E., Welch, C.K., Bell, K.C., Galbreath, K.E., Himes, C. and Allen, J.M., 2014. Natural history collections as emerging resources for innovative education. *BioScience*, 64(8): 725–734. DOI: <http://dx.doi.org/10.1093/biosci/biu096>
- Doherty, P., Guo, Q., Liu, Y., Wieczorek, J. and Doke, J., 2011. Georeferencing incidents from locality descriptions and its applications: A case study from Yosemite National Park Search and Rescue. *Transactions in GIS*, 15: 775–793. DOI: <http://dx.doi.org/10.1111/j.1467-9671.2011.01290.x>
- Ellwood, E.R., Dunckel, B.A., Flemons, P., Guralnick, R., Nelson, G., Newman, G., Newman, S., Paul, D., Riccardi, G., Rios, N., Seltmann, K.C. and Mast, A.R., 2015. Accelerating the digitization of biodiversity research specimens through online public participation. *BioScience*, 65: 383–386. DOI: <http://dx.doi.org/10.1093/biosci/biv005>
- Elwood, S., Goodchild, M.F. and Sui, D.Z., 2011. Researching volunteered Geographic Information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers*, 102: 571–590. DOI: <http://dx.doi.org/10.1080/00045608.2011.595657>
- Environmental Systems Research Institute, 2014. ArcGIS Platform 10.2.
- Eveleigh, A., Jennett, C., Blandford, A., Brohan, P. and Cox, A.L., 2014. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. Association for Computing Machinery, pp. 2985–2994. DOI: <http://dx.doi.org/10.1145/2556288.2557262>
- Fernández, M., Hamilton, H.H. and Kueppers, L.M., 2015. Back to the future: Using historical climate variation to project near-term shifts in habitat suitable for coast redwood. *Global Change Biology*. DOI: <http://dx.doi.org/10.1111/gcb.13027>
- Fleet, C., Kowal, K.C. and Pridal, P., 2012. Georeferencer: Crowdsourced georeferencing for map library collections. *D-Lib Magazine*, 18: 5. DOI: <http://dx.doi.org/10.1045/november2012-fleet>
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69: 211–221. DOI: <http://dx.doi.org/10.1007/s10708-007-9111-y>
- Gotelli, N.J., Chao, A., Colwell, R.K., Hwang, W.H. and Graves, G.R., 2012. Specimen-based modeling, stopping rules, and the extinction of the Ivory-Billed Woodpecker. *Conservation Biology*, 26: 47–56. DOI: <http://dx.doi.org/10.1111/j.1523-1739.2011.01715.x>
- Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Townsend Peterson, A. and Loiselle, B.A., 2008. The influence

- of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology*, 45: 239–247. DOI: <http://dx.doi.org/10.1111/j.1365-2664.2007.01408.x>
- Gregory, I.N. and Ell, P.S., 2007. Historical GIS: Technologies, methodologies, and scholarship. Cambridge University Press. DOI: <http://dx.doi.org/10.1017/CBO9780511493645>
- Guo, Q., Liu, Y. and Wiecek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22: 1067–1090. DOI: <http://dx.doi.org/10.1080/13658810701851420>
- Haklay, M. and Weber, P., 2008. OpenStreetMap: User-generated street maps. *Pervasive Computing, IEEE*, 7: 12–18. DOI: <http://dx.doi.org/10.1109/MPRV.2008.80>
- Hanken, J., 2013. Biodiversity Online: Toward a network integrated biocollections alliance. *BioScience*, 63: 789–790. DOI: <http://dx.doi.org/10.1525/bio.2013.63.10.4>
- Holdsworth, D.W., 2003. Historical geography: New ways of imaging and seeing the past. *Progress in Human Geography*, 27: 486–493. DOI: <http://dx.doi.org/10.1191/0309132503ph440pr>
- Hope, A.G., Waltari, E., Payer, D.C., Cook, J.A., and Talbot, S.L., 2013. Future distribution of tundra refugia in northern Alaska. *Nature Climate Change*, 3: 931–938. DOI: <http://dx.doi.org/10.1038/nclimate1926>
- Hunter, J., Alabri, A. and Ingen, C., 2013. Assessing the quality and trustworthiness of citizen science data. *Concurrency and Computation: Practice and Experience*, 25: 454–466. DOI: <http://dx.doi.org/10.1002/cpe.2923>
- Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T. and Downing, J.A., 2015. Recreational demand for clean water: Evidence from geotagged photographs by visitors to lakes. *Frontiers in Ecology and the Environment*, 13: 76–81. DOI: <http://dx.doi.org/10.1890/140124>
- Kumar, N. and Seitz, S., 2014. Photo recall: Using the Internet to label your photos. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. pp. 785–792. DOI: <http://dx.doi.org/10.1109/CVPRW.2014.120>
- Lamos, V. and Cristianini, N., 2010. Tracking the flu pandemic by monitoring the social web. In *2nd International Workshop on Cognitive Information Processing (CIP)*. pp. 411–416. DOI: <http://dx.doi.org/10.1109/cip.2010.5604088>
- Lauriault, T.P. and Mooney, P., 2014. Crowdsourcing: A geographic approach to public engagement. Paper presented at Internet Politics and Policy Conference. Oxford Internet Institute.
- Liu, Y. and Liu, M., 2015. An online learning approach to improving the quality of crowd-sourcing. Available at <http://web.eecs.umich.edu/~mingyan/pub/Sigmetrics15.pdf>.
- Meier, P., 2012. How crisis mapping saved lives in Haiti. National Geographic Voices. Available at <http://newswatch.nationalgeographic.com/2012/07/02/crisis-mapping-haiti/>.
- Miller, J.S., Porter-Morgan, H.A., Stevens, H., Boom, B., Krupnick, G.A., Acevedo-Rodríguez, P., Fleming, J. and Gensler, M., 2012. Addressing target two of the Global Strategy for Plant Conservation by rapidly identifying plants at risk. *Biodiversity and Conservation*, 21: 1877–1887. DOI: <http://dx.doi.org/10.1007/s10531-012-0285-3>
- Offen, K., 2012. Historical geography II: Digital imaginations. In *Progress in Human Geography*. DOI: <http://dx.doi.org/10.1177/0309132512462807>
- QGIS Development Team, 2014. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at <http://qgis.osgeo.org>.
- R Core Development Team, 2014. R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna, Austria.
- Rivers, M.C., Taylor, L., Brummitt, N.A., Meagher, T.R., Roberts, D.L. and Lughadha, E.N., 2011. How many herbarium specimens are needed to detect threatened species? *Biological Conservation*, 144: 2541–2547. DOI: <http://dx.doi.org/10.1016/j.biocon.2011.07.014>
- Rowe, R.J., 2005. Elevational gradient analyses and the use of historical museum specimens: A cautionary tale. *Journal of Biogeography*, 32: 1883–1897. DOI: <http://dx.doi.org/10.1111/j.1365-2699.2005.01346.x>
- Spyratos, S. and Lutz, M., 2014. Characteristics of citizen-contributed Geographic Information. In Huerta, S. and Granell (Eds.), *Proceedings of the AGILE 2014 International Conference on Geographic Information Science*. Castellón.
- Sui, D. and Goodchild, M., 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, 25: 1737–1748. DOI: <http://dx.doi.org/10.1080/13658816.2011.604636>
- Sui, D., Goodchild, M. and Elwood, S., 2013. Volunteered Geographic Information, the Exaflood, and the growing digital divide. In Sui, D., Elwood, S. and Goodchild, M. (Eds.), *Crowdsourcing Geographic Knowledge*. Springer Netherlands, pp. 1–12. DOI: <http://dx.doi.org/10.1007/978-94-007-4587-2>; http://dx.doi.org/10.1007/978-94-007-4587-2_1
- Taylor, M.L., Yesson, C., Agnew, D.J., Mitchell, R.E. and Rogers, A.D., 2013. Using fisheries by-catch data to predict octocoral habitat suitability around South Georgia. *Journal of Biogeography*, 40: 1688–1701. DOI: <http://dx.doi.org/10.1111/jbi.12122>
- Tingley, M.W. and Beissinger, S.R., 2009. Detecting range shifts from historical species occurrences: New perspectives on old data. *Trends in Ecology and Evolution*, 24: 625–633. DOI: <http://dx.doi.org/10.1016/j.tree.2009.05.009>
- Zhang, M-G., Zhou, Z-K., Chen, W-Y., Slik, J.F., Cannon, C.H. and Raes, N., 2012. Using species distribution modeling to improve conservation and land use planning of Yunnan, China. *Biological Conservation*, 153: 257–264. DOI: <http://dx.doi.org/10.1016/j.biocon.2012.04.023>

How to cite this article: Ellwood, E R, Bart, H L, Jr, Doosey, M H, Jue, D K, Mann, J G, Nelson, G, Rios, N and Mast, A R 2016 Mapping Life – Quality Assessment of Novice vs. Expert Georeferencers. *Citizen Science: Theory and Practice*, 1(1): 4, pp.1–12, DOI: <http://dx.doi.org/10.5334/cstp.30>

Submitted: 13 May 2015 **Accepted:** 18 September 2015 **Published:** 20 May 2016

Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 