

Applying GIS Analysis to Herbarium Georeferencing

By
Michelle Buonopane

Applied Systematics Professional Masters Program
Internship Technical Report

Botany and Plant Pathology
Oregon State University
Corvallis, OR

August 2005

Table of Contents

Project Background.....	1
Vascular Plant Atlas Database Needs.....	3
Materials and Methods.....	6
Results and Discussion.....	7
County and Township/Range Mismatches.....	7
Comparison of ORNHIC and Herbarium Georeferencing.....	7
Creation of Quarter Section and Quarter-Quarter Section LocNames.....	9
Generation of Elevation Information for OrLook LocNames.....	9
Comparison of DEM and Specimen Label Elevations.....	10
Creation of High-Elevation and Low-Elevation Species Lists.....	10
References.....	12
Appendices.....	13
A. Metadata: Creation of Township\Range LocNames Cut by County Lines.....	13
B. Metadata: Comparison of ORNHIC and Herbarium Georeferencing.....	14
C. Differences between OSU Herbarium and Oregon Natural Heritage Information Center (ORNHIC) Georeferencing.....	15
D. Metadata: Creation of Quarter Section and Quarter-Quarter Section LocNames.....	16
E. Metadata: DEM Utilization with Atlas and OrLook Databases.....	18

List of Figures

Figure 1. Public Land Survey System (PLSS).....	3
Figure 2. Example of township/range and location information provided on a Vascular Plant Atlas specimen label.....	4
Figure 3. Oregon Digital Elevation Model.....	6
Figure 4. OSU Herbarium and ORNHIC data points plotted with the boundary of the state of Oregon.....	7
Figure 5. OSU Herbarium and ORNHIC data points plotted with the boundary of the state of Oregon.....	8
Figure 6. Northern section of the state of Oregon along the Columbia River showing two identical specimen records highlighted in blue, one georeferenced by herbarium staff and the other by ORNHIC staff.....	8

List of Tables

Table 1. Average distance in miles between ORNHIC and Herbarium point locations using coordinate information contained in the ORNHIC table (coordinate) and centroids calculated from the ORNHIC polygons (polygon).....	9
Table 2. Examples of the quality control comparison between the Vascular Plant Atlas elevation data and the Digital Elevation Model (DEM) data.....	10
Table 3. Example of a low elevation species list generated from the Oregon Digital Elevation Model data.....	11

Abstract

In March 2005 the Oregon State University Herbarium launched its new, free online Oregon Vascular Plant Atlas for public use. This Atlas allows users to access location information from over 385,000 vascular plant specimen and observation records in the state of Oregon. The potential applications of such spatial information are almost limitless; users can examine the spatial relationships between plants and a variety of environmental variables within plant communities (soils, precipitation, elevation etc.). The maintenance and accuracy of this database is therefore critically important in order to provide the highest quality information to Atlas users. Several projects were identified to increase the accuracy and precision of the Plant Atlas spatial data including: (1) correction of points with mismatched township/range and county information, (2) comparison of Oregon State University (OSU) Herbarium and Oregon Natural Heritage Information Center (ORNHIC) georeferencing results, (3) creation of quarter section and quarter-quarter section location names, and (4) analysis of elevation information using digital elevation models. These projects were completed using geographic information system (GIS) software, as well as Microsoft Access data management software. New location names, or LocNames, were created for township/ranges bisected by county lines, increasing the accuracy of the referenced location. LocNames were also created for specimens with quarter section and quarter-quarter section information. These combined projects resulted in the creation of 1,876,587 new location names. Analysis of the ORNHIC and OSU Herbarium georeferencing results showed the majority of points were similarly georeferenced, however, one-third had some quality issue. Finally, the quality of the elevation information stored in the Vascular Plant Atlas database was shown to include potential data entry conflicts between meters and feet, as well as potential georeferencing problems. The results of these projects will help increase the accuracy of georeferenced specimen locations, as well as provide important information to help herbarium staff identify areas of needed quality control.

1. Project Background

The Oregon State University Herbarium stores plant, fungus, lichen and algal specimens from all over the world, with a special focus on specimens collected in Oregon and the Pacific Northwest. In total, the herbarium holds approximately 405,000 specimens. The information for over 100,000 of the vascular plant specimens has been entered into a database that includes observation records and is accessible and searchable online with over 385,000 records (Oregon Flora Project 2005).

The database includes all of the information contained on the labels of individual specimens. The database is not, however, simply an inventory of the herbarium's vast collection. Each specimen is also georeferenced, meaning latitudinal and longitudinal coordinates are assigned based on the location information on the label. By assigning specific coordinates to each specimen, the location of each specimen can be mapped. The mapping of all the individual plant specimens allows the herbarium staff to create distribution maps for all of Oregon's vascular plants. This mapping project is the Oregon Vascular Plant Atlas, now available online to the public (see <http://www.oregonflora.org/oregonplantatlas.html>). The

atlas combines Oregon maps such as landforms, ecoregions, mean annual precipitation and county lines with the vascular plant taxa the user selects. Thus an atlas user could select a vascular plant of interest and find the locations in Oregon where it has been collected or observed. The user can then combine these locations with one of the Oregon maps listed above. The questions the georeferencing database, or Vascular Plant Atlas, could answer are limitless. Spatial relationships between birds and plants, soils and plants, or elevation and plants are just a small example of the power of the database. The Vascular Plant Atlas has the potential to increase our knowledge about plant distributions and assemblages.

The accuracy of the geographical data becomes incredibly important if it is to be a valuable tool for those interested in its use. Thus the processes involved in georeferencing must minimize errors and maximize accuracy and precision. With an herbarium housing hundreds of thousands of specimens, specimens must be prioritized based on the importance of including them into the Vascular Plant Atlas database. Herbarium staff have made it a priority to georeference at least one specimen of each taxon for each county (Oregon State University Herbarium 2004). At the same time, the ultimate goal is to georeference all the specimens in the herbarium. In order to achieve this, the georeferencing process has to be efficient. An automated georeferencing procedure was designed so that two-thirds of the specimens are georeferenced by computer programs (Mitchell 2005). Specimens with more complicated location information are georeferenced manually by herbarium staff. Quality control programs are continually run on the georeferenced datasets to pick up any data mismatches. For example, a specimen label may list the county in which the plant was collected and provide information about the location in terms of the number of miles from the nearest town. Herbarium staff will assign a location to the plant based on the miles from the town listed. This may put the plant in a county different from that listed on the label. Quality control programs find these county mismatches.

The heart of the automated georeferencing system is the creation of LocNames. LocNames are location names which can be assigned to specimen records, thus eliminating the need to assign specific latitude and longitude coordinates to each specimen. Latitude and longitude information are instead assigned to the LocNames, for example 'Corvallis'. When each LocName is created, an associated "fuzz factor" is assigned. The "fuzz factor" represents the margin of error for a given location represented as a radius in miles (Oregon State University 2004). For example, a specimen can be assigned a specific location but the fuzz factor tells the user the plant is likely to be found within a quarter mile of the mapped point. The OrLook database contains all of the known LocNames and their associated spatial coordinates and fuzz factors. Automated computer programs search the entered location information (indicated on the specimen label) for a given record and automatically match it to known LocNames. For more complicated specimen label location information, herbarium staff manually search for the best LocName. For those specimens with no matching LocName, new LocNames are created. The greater the number of LocNames in the OrLook database, the greater the number of specimens which can be georeferenced automatically by the computer programs.

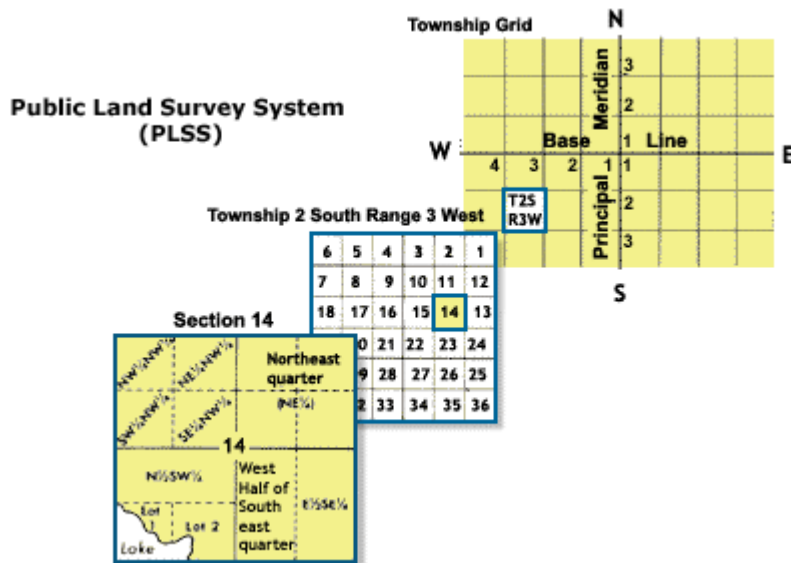
Vascular Plant Atlas Database Needs

Herbarium staff are continually looking for new ways of increasing the accuracy of the georeferencing process, as well as identifying possible ways of verifying the quality of the data the Atlas already utilizes. The data are valuable only if the georeferencing process accurately interprets the location information contained on specimen labels. Herbarium data managers identified several areas in which the quality of this process could be improved.

1. Correcting for county and township/range mismatches.

One of the problems in the current Vascular Plant Atlas database is county mismatches with the specimen's township and range information. Township and range information is part of the Public Land Survey System (PLSS) established with the Land Ordinance Act of 1785 and the Northwest Ordinance Act of 1787 (National Atlas of the United States 2005). The PLSS divides federal land into rectangular sections, originally designed to facilitate transfer of mostly western lands to private ownership over time. Townships are 36 square miles in size, bounded by township lines to the north and south, and by range lines to the east and west. Townships are divided into thirty-six one-square-mile sections comprising roughly 640 acres. Sections can be further divided into quarter sections or quarter-quarter sections (Figure 1).

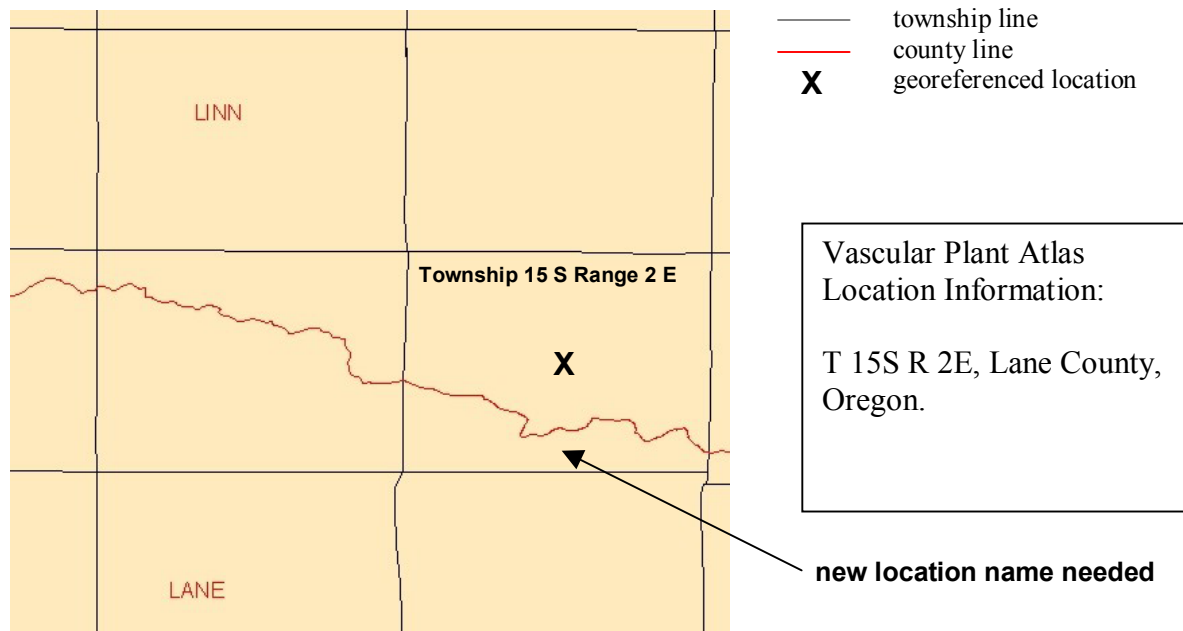
Figure 1. Public Land Survey System (PLSS). Reprinted from www.nationalatlas.gov.



In some cases, only the township and range are indicated on a vascular plant specimen label. The computer program selects the coordinates in the center of the township as the point location for a given specimen. Often, however, the township and range may cross more than one county; therefore a specimen may be assigned a point that does not agree with the county indicated on the specimen label (Figure 2). This results in a county mismatch. These records must be corrected so the point location is moved to the center

of that portion of the township/range within the county indicated on the specimen label. New LocNames are needed for each township/range within each county.

Figure 2. Example of township/range and location information provided on a Vascular Plant Atlas specimen label. Current mapping of this specimen would place the plant record in the wrong county.



2. Comparing the georeferencing of specimens contained in the herbarium database and those same specimens georeferenced by the Oregon Natural Heritage Information Center.

The Oregon Natural Heritage Information Center (ORNHIC) also manages a database of plant, animal and plant community locations. This database was shared with the OSU Herbarium. The ORNHIC database contained some plant specimen records already included in the herbarium database. These records were excluded when the two databases were merged, so the same specimens were not recorded twice. These 490 duplicate records present an opportunity to examine the differences in assigned spatial coordinates between the ORNHIC staff methodology and that of the herbarium staff. Thus the location of the same specimen could be mapped twice, from the ORNHIC database, and the OSU herbarium database, to measure the distance between the two point locations, and give the staff of both organizations an understanding of how their georeferencing procedures differ. This will be especially useful to herbarium staff by potentially increasing their understanding of the relative accuracy of the specimens ORNHIC georeferenced, and simply providing a quality-control method to identify problem records.

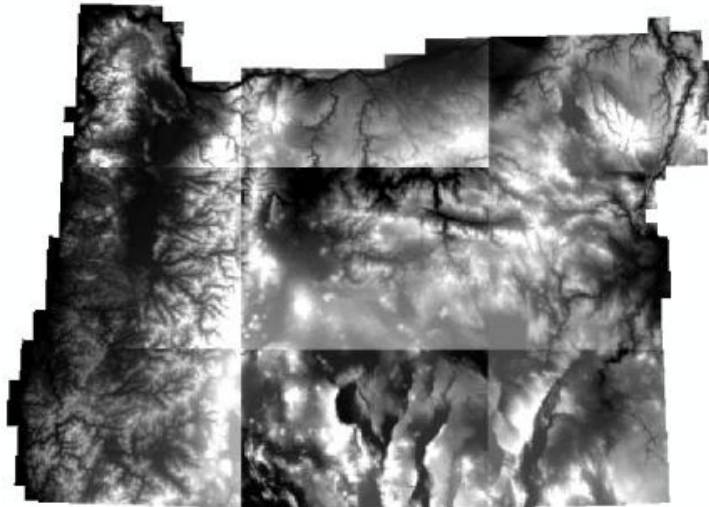
3. Increasing the accuracy of point locations by including quarter section and quarter-quarter section information within township/range information.

Some specimens' labels contain township, range, section and quarter section (or quarter-quarter section) PLSS information (see PLSS description above). To date, coordinates have only been assigned based on the section level because no quarter or quarter-quarter section LocNames have been created in OrLook. New LocNames can be created to assign a point location based on the more accurate scale of the quarter section (or quarter-quarter section).

4. Using digital elevation models to generate elevation information for OrLook LocNames, and allow for quality control of specimen and observation label elevations.

Only a small portion of the OrLook LocName records have associated elevation information. Digital elevation models (DEMs) contain elevation information which can be compared to known Atlas records or to generate elevations for those LocNames with missing elevation information. DEMs are a type of GIS layer utilizing a raster data format. Raster datasets utilize a grid of cells to create a visual image. Each cell, or pixel, displays a single attribute, in this case, elevation. DEMs appear as a patchwork of cells with lighter cells indicating higher elevations and darker cells indicating lower elevations (Figure 3). DEMs make it possible to assign elevations to any known coordinates. Thus, elevations can be generated for all of the OrLook LocNames, including the newly created quarter section LocNames. In addition, DEMs can be used to assess the accuracy of known elevations in the Vascular Plant Atlas database. This allows for two quality control processes: (1) label elevation values can be compared with DEM values to identify problems with incorrect unit assignment (feet versus meters) or data entry errors, and (2) comparison of label and DEM elevations can validate the accuracy of the georeferencing process. For example, one can ask: if a plant was collected at a given location with the elevation recorded on the label, does the georeferencing process place the specimen at an elevation significantly different from that recorded in the field? Finally, lists of plant species restricted to high elevations and low elevations can be generated using DEM or label elevation information.

Figure 3. Oregon Digital Elevation Model. The pieces of the DEM appear as a patchwork because each section is imported separately. Each pixel represents an elevation in feet.



Each of these projects will ultimately allow herbarium staff to identify problem areas requiring greater quality control work, or increase the available LocNames allowing more automated and accurate georeferencing.

2. Materials and Methods

All of the outlined projects required the use of spatial data—information which identifies where objects are located on the Earth’s surface. Geographic Information Systems (GIS) are software programs designed specifically to manage and analyze spatial information. GIS links spatial coordinates to relevant information such as an Oregon mountaintop and a plant species name. ArcGIS 8.3 was the software used in the herbarium (Crosier *et al.* 2002). ArcGIS allows the user to capture, store, query, analyze and display spatial data.

All Oregon spatial layers were downloaded from the Oregon Geospatial Data Clearinghouse (OGDC) including PLSS, county line, state boundary and DEM layers, and were projected in Oregon Geographic Information Council Lambert projection (<http://www.gis.state.or.us/>). PLSS layers were checked for errors before use, and any errors detected were corrected by comparison with topographic maps. ArcView 3.3 was utilized for the creation of the quarter and quarter-quarter section layers in order to employ an avenue script designed to automatically divide PLSS sections into quarters. A Visual Basic script, GridSpot, was downloaded from the ESRI website to pull values from a raster dataset (DEM) and populate an elevation field in the corresponding point layer.

Often the outlined projects required a significant amount of data manipulation and summarization after spatial analysis was conducted in ArcGIS. Microsoft Access 2003 was

used for all data management needs. Means, standard errors and standard deviations for the elevation comparisons were calculated in SAS 8.2 (SAS 1989).

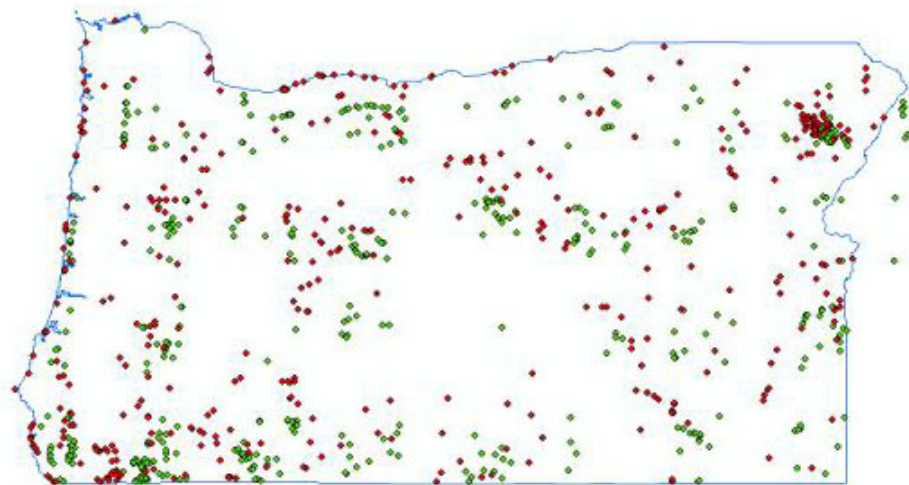
All work was documented and stored with the Oregon Vascular Plant Atlas database manager. For specific details about an individual project's protocol see Appendices A-E.

3. Results and Discussion

County and Township/Range Mismatches- New LocNames were created for township/ranges cut by county lines. Five-hundred and sixty-six township/ranges (out of 2,889) were bisected by a county line at least one time, and 1,173 new LocNames were created. Currently the Vascular Plant Atlas contains 216 records with mismatched township/range and county information; these can now be georeferenced with the new LocNames. Close to three thousand additional records with only township/range information can also be georeferenced. Automated computer programs can now assign LocNames to specimens with township/range information for multiple counties. These new LocNames also allow herbarium staff to manually assign LocNames for more complicated specimen label information. Both accuracy and efficiency have been increased with the creation of these LocNames.

Comparison of ORNHIC and Herbarium Georeferencing- Distances between an individual plant specimen georeferenced by both ORNHIC and herbarium staff were first calculated using the latitude and longitude values listed in the ORNHIC tables. Ninety-eight percent of the points were south and east of the herbarium points (Figure 4). This spatial mapping allowed us to see a consistent problem with the data points. Further investigation showed that the latitude and longitude units in the ORNHIC table were listed in degrees, minutes, seconds instead of decimal degrees. The spatial information was listed as a string of numbers creating the confusion.

Figure 4. OSU Herbarium and ORNHIC data points plotted with the boundary of the state of Oregon. Red dots represent herbarium points and green dots represent ORNHIC points taken from the table latitude and longitude coordinates. Note the green dots are south and east of the red dots.



Distances between the points were then recalculated using latitude and longitude values taken as the center of the ORNHIC polygons (Figure 5). Two-thirds of the ORNHIC points fell within the fuzz surrounding the herbarium point (Appendix C). Thirty-one percent of the ORNHIC points fell outside of the fuzz but within the same county. Thus the majority of the ORNHIC points corresponded well with the spatial location of the herbarium points. Three percent of the points however, fell outside of the fuzz and outside of the county (Figure 6). Six of the ORNHIC points fell greater than 25 miles from the herbarium point and one outlier was 176 miles from the associated herbarium point; these were identified as points requiring further investigation for quality control purposes. Table 1 lists the results of the distance analysis.

Figure 5. OSU Herbarium and ORNHIC data points plotted with the boundary of the state of Oregon. Red dots represent herbarium points and green polygons represent ORNHIC specimen records. Note the red dots are now usually in the center of the polygons.

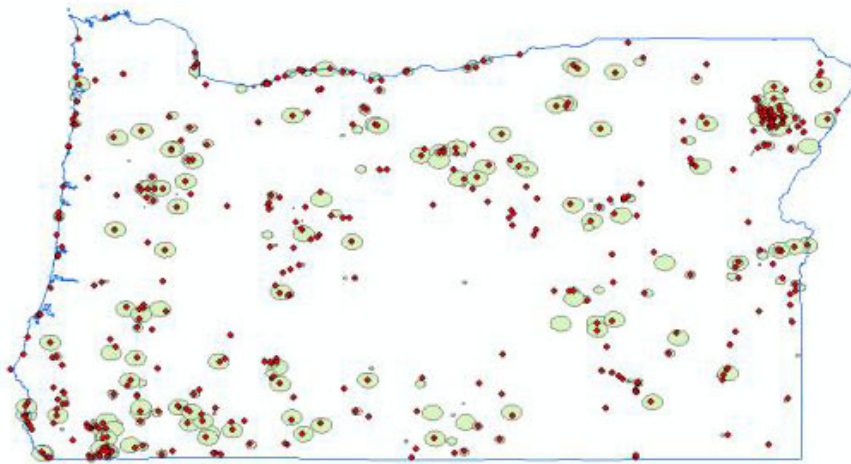
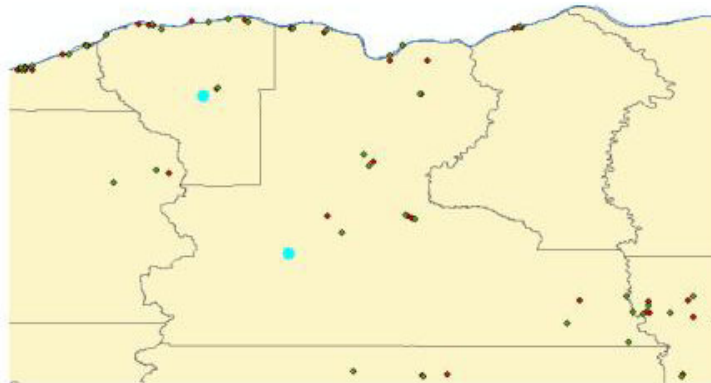


Figure 6. Northern section of the state of Oregon along the Columbia River showing two identical specimen records highlighted in blue, one georeferenced by herbarium staff and the other by ORNHIC staff. The two points fall in different counties. Red dots represent herbarium points and green dots represent ORNHIC points.



Identification of these problem records allowed the herbarium staff to verify the quality of the data points. These types of quality control checks are important sources of information for a database of this size. The points will also be examined to determine if there is a systematic explanation for the discrepancies.

Table 1. Average distance in miles between ORNHIC and Herbarium point locations using coordinate information contained in the ORNHIC table (coordinate) and centroids calculated from the ORNHIC polygons (polygon).

	avg	se	sd	max	min
polygon	2.24	± 0.41	9.08	175.95	0.009
polygon w/out outlier	1.88	± 0.206	43.46		
coordinate	17.63	± 0.47	10.47	176.17	1.7
coordinate w/out outlier	17.31	± 0.35	7.68		

Creation of Quarter Section and Quarter-Quarter Section LocNames- There are 97,943 sections in Oregon. Each of these sections was divided automatically by an ArcView script into four quarter sections and sixteen quarter-quarter sections. Over 375,000 new quarter section LocNames and greater than 1,500,000 new quarter-quarter section LocNames were created. Not all sections could be subdivided. Sections surrounding lakes and along the state boundary (coastal and inland) are often of irregular shape (and thus not rectangular) and usually represent only a portion of a true section (these may only contain two quarter sections or fewer). These irregular sections were automatically divided into four equal pieces by the computer program and could not be used. In addition, some sections could not be subdivided even if appearing rectangular in outline if the number of vertices in the polygon exceeded that allowed by the program. The new LocNames created can now be utilized to assign more accurate spatial coordinates to 7,774 records in the Vascular Plant Atlas database. In addition, these new LocNames will also be used for georeferencing specimens from other herbaria.

Generation of Elevation Information for OrLook LocNames- OrLook LocNames with a fuzz factor less than or equal to a quarter mile (804 m) were used to generate elevation information. Records with a larger fuzz factor would have a margin of error too large to be useful for elevation generation. For example, records with a fuzz factor of six would have a three mile radius. The elevation at the center of the fuzz where the record would be mapped could be very different from the elevation of a point on the perimeter of the three-mile radius circle surrounding it. These records were therefore excluded. Close to 143,000 LocNames were updated with elevation information using Oregon DEMs. Many of these LocNames were already assigned elevation information, especially those which were generated from a United States Geologic Survey (USGS) placenames dataset. It was noted that some of these elevations differed significantly from the DEM generated elevations. These were highlighted for further investigation by herbarium staff. Some of these potential data errors could be a result of a difference in units (meters versus feet), but further work is needed to determine what action is necessary to resolve the discrepancies.

Comparison of DEM and Specimen Label Elevations- Elevations were generated for all of the Vascular Plant Atlas database records, and only those with a fuzz code equaling a quarter mile (804 m) or less were further investigated. Those specimens already assigned specimen label-based elevation data were compared with DEM elevation calculations. The lowest DEM value was 1 foot; thus any points falling within Oregon with a label elevation of zero were not compared. Elevation information from specimen labels was recorded for 94,646 records.

The absolute value of the difference between the DEM elevation and the Vascular Plant Atlas label elevation was calculated for each record. In case the entered Vascular Plant Atlas elevation was actually in feet and not meters, potential unit errors were examined. To assess how meaningful the differences were, the difference was displayed as the percentage of the horizontal fuzz factor distance (PercentofFuzz). For example, if the Vascular Plant Atlas label elevation value was 1100 m and the DEM value was 1121 m, then the difference would be five percent of an 402 m horizontal fuzz factor (Table 2).

Table 2. Examples of the quality control comparison between the Vascular Plant Atlas elevation data and the Digital Elevation Model (DEM) data. Table displays accession number, fuzz factor distance in meters (FuzzDist), plant taxon six-letter acronym, Atlas elevation data as originally entered and converted to meters in case the table value was actually in feet (AdjElev), DEM elevation in meters, absolute value of the calculated difference between the Atlas and DEM meter values, percent of fuzz distance (%Fuzz) and unit error determination (UnitError). Shaded areas highlight records where the adjusted value corrected the calculated difference.

Accession#	FuzzDist	Acronym	Atlas		DEM		Diff	%Fuzz	UnitError
			Elev (m)	AdjElev (m)	Elev (m)				
ORE89242	402	ASTOCC	1100	335	1122	22	5		
OSC193668	804	THAOCC	334	102	455	121	15		
ORE80427	804	CASPRU	1803	550	1520	283	35		
ORE105710	804	SELSCO	1555	474	877	678	84		
ORE81192	402	MIMCUS	1190	363	432	758	189	yes	
OSC194891	804	GNASTR	3180	969	963	2217	276	yes	

Eighty percent of the records showed a good match between the DEM elevation and the label elevation (PercentofFuzz \leq 25) and did not require further quality control examination. Eighteen percent of the records were identified as needing further investigation, those where the 'PercentofFuzz' was greater than 25%. While this amounted to over 17,000 records, these were linked to only 3,628 locations. Thus the locations alone would be the only records requiring quality control work by herbarium staff. 'PercentofFuzz' values greater than 75% were examined individually for potential problems between meters and feet. Close to 1000 records were identified as likely unit errors. The analysis of elevational differences revealed an important and unknown area of work requiring quality control. This was an important tool in identifying potential problems.

Creation of High-Elevation and Low-Elevation Species Lists- Lists of plant taxa restricted to elevations greater than 2500 m and less than 50 m were generated based on the Vascular Plant Atlas label data and the DEM elevations (Table 3). The generated lists noted the number of records for each taxon included in the elevational range summaries allowing the list user to understand the relative strength of the data. This is especially important for identifying cases where only one record exists for a given taxon. These lists represent only an initial capability of the DEM dataset. As the plant Vascular Plant Atlas database grows, the value of these lists will only be strengthened as more records are included in the range lists. These lists also represent an example of the potential for generating species lists based on county boundaries, ecoregions, rainfall patterns and more.

Table 3. Example of a low elevation species list generated from the Oregon Digital Elevation Model data. Species listed were encountered at elevations below 50 m. Count indicates the number of specimens/observations used to generate the list. MaxElev = maximum elevation. AvgElev = average elevation.

Acronym	Taxon	Count	MaxElev	AvgElev
CAKEDE	<i>Cakile edentula</i>	27	42.06	7.80
CAKEDEEDE	<i>Cakile edentula</i> var. <i>edentula</i>	10	42.06	14.78
CARLYN	<i>Carex lyngbyei</i>	30	42.06	10.96
CASAMBAMB	<i>Castilleja ambigua</i> var. <i>ambigua</i>	22	28.04	9.84
ELYMOL	<i>Elymus mollis</i>	21	27.13	11.13
EPIFRA	<i>Epilobium franciscanum</i>	11	32.92	23.89
HONPEPMAJ	<i>Honckenya peploides</i> var. <i>major</i>	13	17.07	10.29
JUNBRE2	<i>Juncus breweri</i>	24	49.99	15.95
JUNCAN	<i>Juncus canadensis</i>	10	32.92	16.34
JUNFALSIT	<i>Juncus falcatus</i> var. <i>sitchensis</i>	11	49.99	19.01
LATJAP	<i>Lathyrus japonicus</i>	16	25.91	11.91
LATPALPAL	<i>Lathyrus palustris</i> var. <i>palustris</i>	13	24.08	6.66
LYSTER	<i>Lysimachia terrestris</i>	15	28.04	14.71
SPEMACMAC	<i>Spergularia macrotheca</i> var. <i>macrotheca</i>	11	17.07	7.07

Utilization of GIS technology has proven to be an extremely effective way to increase the accuracy of the Vascular Plant Atlas database. It offers numerous opportunities to not only maximize the value of the data stored within the Atlas, but to also provide important quality control checks to ensure the results of the data analyses are truly meaningful.

4. References

- Crosier, S., Booth, B. and A. Mitchell. 2002. Getting started with ArcGis: GIS by ESRI. Environmental Systems Research Institute: Redlands, CA.
- Mitchell, Katie. 2005. From data to dots. Oregon Flora Newsletter 10(2&3): 11.
- National Atlas of the United States. 2005. The public land survey system (PLSS). Accessed July 23, 2005 from http://nationalatlas.gov/articles/boundaries/a_plss.html.
- Oregon Flora Project. 2005. Oregon Plant Atlas. Accessed June 10, 2005 from <http://www.oregonflora.org/oregonplantatlas.html>.
- Oregon State University Herbarium. 2004. Specimen georeferencing protocol. Version 2.2
- SAS Institute Inc. 1989. SAS/STAT Users Guide, Version 6, Fourth Edition, Volume 1. SAS Institute Inc.: Cary, NC.
- National Atlas of the United States. 2005. The public land survey system (PLSS). Accessed July 23, 2005 from http://nationalatlas.gov/articles/boundaries/a_plss.html.

Appendix A: Metadata: Creation of Township\Range LocNames Cut by County Lines

Creation of TR LocNames (and centroids) where TR Cut by County Line:

(File Location:

T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\townshprng\TRonly\Temp\Final_TRC
nty.dbf)

- 1) Chose to use townships_100K_geo.shp layer because it was a larger scale than townships_2000K_geo.shp (although TRS layer used 2000K).
- 2) Checked for mistakes in 100K file.
Compared 100K layer with 2000K by creating a layer of 100K centroids and joined them to the 2000K polygons (points that fall within polygon). This was input into Access to find 100K TR labels that did not correspond with 2000K TR labels. The mismatches were then mapped in ArcMap. (Note: these did not include centroids which fell into the ocean or into Idaho). There were four digitizing errors and 8 TR misname errors. These were corrected by comparing with TRs on topomaps. The polygons with the same TRs (those separated by lakes or rivers) were then dissolved so centroids could be recalculated on the total area of the TR.
- 3) TRs were then selected that were cut by the outline of the counties (orcnty24_geo883.shp layer used).
- 4) Created new shape file of only selected records.
- 5) Intersected the selected TRs with the county lines.
Some TRs were cut by a single county line more than once. I consulted with Katie and we chose to select the polygon with the larger area for centroid calculation. These TRs were given 'Tc' as an origin. Smallest polygon included in creation of new LocNames was 0.00001 decimal degrees. As a result, we omitted a large number of slivers created when the edges of TR polygons did not line up exactly with County polygons. (Three TRs were cut by county lines 4 times; 35 TRs were cut by county lines 3 times; 528 TRs were cut by county lines 2 times.)
- 6) New LocNames were then created for OrLook (file= LocName_TRCutByCnty.dbf).

Creation of TR Locnames for all TRs:

- 1) Created centroid layer based on townships_100K_geo.shp corrected layer.
- 2) Joined centroids to county polygons (points that fall within polygons).
- 3) Imported layer into Access.
- 4) Created new LocNames.
- 5) Found TRs which were cut by counties by joining to LocName_TRCutByCnty.dbf.
Created new LocNames and County names for Unknown County records.
- 6) Joined back to LocName_TRs. LocNames and Counties names were updated.

This created two files for updating OrLook: LocName_TRCutByCnty.dbf and LocName_TRs.dbf

Appendix B: Metadata: Comparison of ORNHIC and Herbarium Georeferencing

Associated files:

T:\Herbarium\OFPDB\OFPDatabase\GIS\GIS files\Temp_ORNHIC\Report on Distances.doc

T:\Herbarium\OFPDB\OFPDatabase\GIS\GIS files\Temp_ORNHIC\Herb_ORNHIC_centroids.xls

T:\Herbarium\OFPDB\OFPDatabase\GIS\GIS files\Temp_ORNHIC\ORNHIC_Specimens.mdb, table = Distances_centroids

1. ORNHIC polygon data (file=ORNHIC_Specimens1.dbf) were imported into ArcMap and centroids were calculated from the polygon data. (Note: initially the latitude and longitude coordinates were used to generate an ORNHIC point layer but 96% of these points were all consistently SE of the herbarium points and huge distance errors resulted from using these points. Centroids were then calculated as better point comparisons.)
2. The ORNHIC and Herbarium point layers (file=Specimens_Herbarium_Georef.dbf) were joined on accession number in ArcMap, and imported into access, and calculations were done in excel. Distances between points based on accession number were calculated based on the following equation:

$$D = 60 \cos^{-1} [\sin L_1 \sin L_2 + \cos L_1 \cos L_2 \cos(\lambda_1 - \lambda_2)]$$

D = distance (nautical miles), where 1 nautical mile = 1.151 stat. miles

L₁ = Original latitude (radians)

L₂ = Destination latitude (radians)

λ₁ = Original longitude (radians)

λ₂ = Destination longitude (radians)

Note: ArcCos returns result in radians and therefore you have to convert radians to degrees before multiplying by 60 (degrees = (radians/Pi)*180).

3. ORNHIC points were identified as problematic if they fell outside the horizontal fuzz distance of the herbarium point. (See Report on distances.doc)
4. ORNHIC points were also identified as potentially problematic if they fell inside the herbarium horizontal fuzz but outside of the specimen county.
5. See file T:\Herbarium\OFPDB\OFPDatabase\GIS\GIS files\Temp_ORNHIC\Herb_ORNHIC_centroids.xls for detailed distance data. See file T:\Herbarium\OFPDB\OFPDatabase\GIS\GIS files\Temp_ORNHIC\ORNHIC_Specimens.mdb, table = Distances_centroids for final table on distances and counties.

Appendix C: Differences between OSU Herbarium and Oregon Natural Heritage Information Center (ORNHIC) Georeferencing

When the reference location was calculated using the center of the ORNHIC polygons, the following differences were detected when compared with the herbarium georeferencing results:

66.5% of ORNHIC points fall within the herbarium “fuzz”¹
31% of ORNHIC points fall outside of the fuzz but within the same county
3 % of ORNHIC points fall outside of the fuzz AND outside the county
6 ORNHIC points fall greater than 25 miles from the herbarium point²

Distance between points (miles):	avg = 2.24	without outlier ³ :	avg = 1.88
	stdev = 9.08		stdev = 4.56
	sterr = 0.41		sterr = 0.206
	max = 175.95		
	min = 0.009		

When the reference location was based on the latitude and longitude provided in the ORNHIC data table the following differences were detected when compared with the herbarium georeferencing results:

2.5% of ORNHIC points fall within herbarium “fuzz”
2.7% of ORNHIC points fall outside of the Oregon state boundary
68% of ORNHIC points fall outside of the fuzz but within the same county
26% of ORNHIC points fall outside of the fuzz AND outside the county
19% of all ORNHIC points fall greater than 25 miles from the herbarium point

Distance between points (miles):	avg = 17.63	without outlier ³ :	avg = 17.31
	stdev = 10.47		stdev = 7.68
	sterr = 0.47		sterr = 0.35
	max = 176.17		
	min = 1.7		

Note: It was later discovered that the latitude and longitude numbers provided in the ORNHIC table were not listed in decimal degrees.

¹ A fuzz level is the error associated with each georeferenced specimen, indicated as a radius in miles surrounding a georeferenced point.

² The six ORNHIC points greater than 25 miles, listed by accession number: ORE81182, OSC109408, OSC11753, OSC143925, WILLU13998, WILLU29400

³ outlier: one ORNHIC point falling 176 miles from the herbarium point

Appendix D: Metadata: Creation of Quarter Section and Quarter-Quarter Section LocNames

1. Corrected state plss layers. Files: plss_no_dlc.shp. Corrected the following polygons:
 - a. 9S 13E 0 changed to 9S 13E 31
 - b. 39S 7E 0 changed to 39S 7E 27Some polygons (one section) were divided into two polygons, with one polygon having a section number of 0. I removed remaining polygons where section = 0. Then I corrected those polygons that were now missing half of the section by correcting the vertices.
 - c. 17S 3W 25
 - d. 3N 12E 33 and 3N 12E 31
 - e. 28S 15W 36This file was named plss_no_dlc_edited_edvertic.shp
2. Plss_no_dlc_edited_edvertic.shp was used to create the quarter and quarter-quarter section layers in ArcView 3.3.
3. An avenue script was used to create the sections. Script file name = quartersections.apr.
4. I ran the quarter section program with a distance of 1000 (units unknown) to calculate quarter sections for polygons with more than 4 vertices. I ran the quarter section program with a distance of 500 (units unknown) to calculate quarter-quarter sections for polygons with more than 4 vertices.
 - a. Quarter Section Layer: 95 polygons were not divided; 529 had more than one vertex per side and were NOT divided; 58511 polygons had more than 1 node per side and WERE divided; out of a total of 97883. NOTE: sections that are not rectangular (i.e. triangles) are still divided the same way and should not be used.
 - b. Quarter-Quarter Section Layer: 95 polygons were not divided; 692 had more than one vertex per side and were NOT divided; 14688 polygons had more than 1 node per side and WERE divided; out of a total of 97883.
5. New files were created: plss_no_dlc_qrtsec.shp and plss_no_dlc_qrtqrtsec.shp
6. File projection was then defined by importing the original projection from plss_no_dlc_edited_edvertic.shp.
7. Files were then reprojected to NAD 83 and named plss_no_dlc_qrtsec_geo.shp and plss_no_dlc_qrtqrtsec_geo.shp.
8. The projected shapefile was then used to calculate centroids for each quarter-section and quarter-quarter section. A new point layer was created from the centroids plss_no_dlc_qrtsec_xy.shp and plss_no_dlc_qrtqrtsec_xy.shp.
9. The point layers were then clipped to the state boundary layer: state_boundary_24_geo_NoWater_FewAttributes.shp.
10. In order to remove the points located along the boundary of the state (which were incorrectly generated by the avenue script), a 1 mile buffer was created along the state boundary. Only centroids outside of this 1 mile buffer (within the state) were included.

11. Polygons that were smaller than 20,000,000 in area were removed because the avenue script did not correctly divide partial sections associated with lakes and streams (see below).
12. Polygons that were not subdivided by the avenue script were also excluded from the dataset.
13. The file with final LocNames and xy coordinates can be found:
T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\townshprng\plss_no_dlc\Quarters\QuarterSections_LocName.mdb; table name = plss_no_dlc_qrtsec_FINAL and plss_no_dlc_qrtqsec_FINAL

Problems with this methodology:

- 1) State boundaries are sometimes irregular resulting in irregularly shaped section polygons. The program divides these polygons incorrectly and therefore no LocNames are created for these boundary quarter-sections.
- 2) Boundaries of lakes within the state are also often irregular. Section polygons along lakeshores are therefore often irregular. The program divides these polygons incorrectly and therefore no LocNames are created for these lake boundary quarter-sections.
- 3) Some of the Oregon section polygons are larger, and some are smaller than the average 640 acre size. The avenue script assumes all polygons should be subdivided into regular quarters regardless of size.

Appendix E: Metadata: DEM Utilization with Atlas and OrLook Databases

Associated Files:

T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\Atlas_DEM.mdb

T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\Atlas_elevations.mdb

T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\OrLook_elevations.mdb

T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\tnwshprng\plss_no_dlc\Quarters\QuarterSections_LocName_elev.mdb

DEM and Atlas Data Known Elevation QC

1. Elevations of Atlas records were calculated in ArcGIS with 10 meter digital elevation models (DEM) using layers downloaded from <http://buccaneer.geo.orst.edu/dem/>. (The USGS developed a national elevation dataset (NED). The NED is a seamless mosaic of best available elevation data. The 7.5 minute elevation data for the conterminous US are primarily initial source data. The 10 m dems for Oregon were merged whenever possible in their native UTM projection. The resulting file was projected into the Oregon Lambert projection and clipped into 9 separate grids for the state of Oregon.)
2. Atlas records and new Specimen records not yet entered into the Atlas database were used..
3. Elevations were assigned using GRIDSPOT visual basic script downloaded from the ESRI website (gridspot.frx). GRIDSPOT takes the elevation from the DEM layer for a corresponding point in a shapefile and populates a new field in the shapefile.
4. In situations where DEM sections overlapped, two elevations were calculated for a given specimen/observation record. If these values differed, the average of the two values was used. *Records that fell outside of a given DEM section were assigned zeros and thus "false zero" values were deleted from the dataset.*
5. For records where an elevation was entered in the database from specimen/observation information, a quality control program was run to examine the difference between the Atlas elevation values and the DEM elevation values. The Access database can be found:
T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\Atlas_DEM, table = Atlas_Elev_Qc.
 - a. The absolute values of the difference between the DEM elevations (field = DEM_METER) and the Atlas elevations (field = ELEVMMIN) were calculated (field = Diff_DEM_m). In case the entered Atlas elevation was actually in feet and not meters, the Atlas elevations (=ELEVMMIN) were converted to meters (=ELEVMMIN_con). The difference was again calculated but this time for the converted elevations and the DEM values (=Diff_DEM_con). To assess how meaningful the differences may be, the difference was displayed as the percentage of the horizontal fuzz distance [$\text{Percentoffuzz} = (\text{Diff_DEM_m}) / (\text{Fuzz_dist_M}) * 100$]. Percentoffuzz values greater than 75% were examined by hand. Any situations where the difference could be attributed to an error in units were assigned a 'UnitsError = yes' label. Those which could be possible unit errors were assigned a

'UnitsError = ?' label. Those points where the DEM value was zero were coded as "out".

- b. In situations where the Atlas dataset had both minimum and maximum elevations (ranges), the absolute value of the difference between the DEM elevation (=DEM_METER) and the Atlas minimum (=ELEVMMIN) AND Atlas maximum (ELEVMMAX) was calculated (fields = Range_diff_min and Range_diff_max, respectively). PercentofFuzz was calculated as before using the minimum elevation difference. UnitsError was also calculated as above.

6. The QC was done ONLY on those records with a fuzz less than or equal to 4 and with a specimen/observation label elevation. But elevations were calculated from DEMs for all records.

percent of fuzz	# of records	# of sites	units error	# of records	# of sites
>100	1573	581	yes	985	424
75 - 99	1267	236	?	367	19
50 - 74	3704	595	out	10	2
25 - 49	11177	2226	total	1362	445
0 - 25	57814	15274			
total	75535	18895			

total # of records = 94646*

of records with no QC problems = 19111

percentage of records with problems = 18**

*fuzz less than or equal to 4 or L4

**'problem record' defined as those with percent of fuzz >25

fuzz link	percent of fuzz	# of records	# of sites/specimens
1	>100	7	7
	75 - 99	0	0
	50 - 74	88	14
	25 - 49	67	6
	0 - 25	136	10
2	>100	605	16
	75 - 99	266	10
	50 - 74	898	24
	25 - 49	612	32
	0 - 25	2516	91
3	>100	99	90
	75 - 99	303	43
	50 - 74	669	54
	25 - 49	1754	192
	0 - 25	7983	729
4	>100	862	468
	75 - 99	698	183
	50 - 74	2049	503
	25 - 49	8744	1996
	0 - 25	47179	14444
	total	75535	18912

Atlas Data Elevation by Taxon

1. Mean, maximum and minimum elevations by taxon were calculated with the Digital Elevation Model (DEM) data (as assigned above). Standard deviations and standard errors were also calculated. File=
T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\Atlas_elevations.mdb
Table = Sppmean_DEM. The same numbers were generated for only those records where problems were not identified above (i.e. PercentofFuzz = 0). Filename =
Range_mean_nullfuzz.xls
2. Mean, maximum and minimum elevations by taxon were calculated using the label elevation information for those specimens/observations with such data. Standard deviations and standard errors were also calculated. Table = Sppmean_LABEL.
3. Taxa restricted to elevations greater than 2500 meters were determined using both DEM data and label data (Table name: SPP>2500 and SPP>2500_LABEL, respectively).
4. Taxa restricted to elevations less than 50 meters were determined using both DEM data and label data (Table name: SPP<50 and SPP<50_LABEL, respectively).

OrLook Elevation Generation Using Oregon DEMs

Completed: June 2005 By: Michelle Buonopane

1. Records with LocNames with a fuzz equal to or less than 4 (or L4) were selected from the OrLook database.
2. A point layer was created from these records and imported into ArcGIS.
3. Elevations of OrLook LocNames were calculated in ArcGIS with 10 meter digital elevation models (DEM) using layers downloaded from <http://buccaneer.geo.orst.edu/dem/>. (The USGS developed a national elevation dataset (NED). The NED is a seamless mosaic of best available elevation data. The 7.5 minute elevation data for the conterminous US are primarily initial source data. The 10 m dems for Oregon were merged whenever possible in there native UTM projection. The resulting file was projected into the Oregon Lambert projection and clipped into 9 separate grids for the state of Oregon.) The DEM layer is divided into 9 sections.
4. Elevations were assigned using GRIDSPOT visual basic script downloaded from the ESRI website (gridspot.frx). GRIDSPOT takes the elevation from the DEM layer for a corresponding point in a shapefile and populates a new field in the shapefile.
5. The new table was imported into Access where records with an elevation of zero were deleted. (No elevations of zero are assigned in GRIDSPOT unless a point falls OUTSIDE of the DEM layer. DEM elevation values are always greater than 0. Elevations are converted to meters in Access. The file is located:
T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\DEM\OrLook_elevations.mdb
, table = OrLook_ElevTable
6. DEM layers do overlap and two different elevations can be created for the same point. In these cases the average of the two elevations were taken and assigned for a given point and duplicates deleted.

7. Some OrLook records contain elevation information already (field = ElevMLook). There are some QC issues with these values. In some cases there may be elevations in the "ElevMLook" field where units = feet instead of meters. Some elevations also differ largely from the DEM value. Some querying may be necessary to resolve these QC issues.

Quarter Section and Quarter-Quarter Section Elevation Generation Using DEMs

Completed: June 2005 By: Michelle Buonopane

1. The same procedure was used as outlined above for the quarter section and quarter-quarter section LocNames datasets.
2. Twenty-six quarter section LocNames could not be assigned elevation. One hundred and sixty-two quarter-quarter section LocNames could not be assigned elevation.
3. The files can be found in:
T:\Herbarium\OFPDB\OFPDatabase\GIS\Flora_layers\townshprng\plss_no_dlc\Quarters\QuarterSections_LocName.mdb; table name = plss_no_dlc_qrtsec_elev and
plss_no_dlc_qrtqrtsec_elev