

CHITTENDEN COUNTY DWELLING UNITS GIS DATABASE

To understand the nature of housing in Chittenden County, it is necessary to inventory our existing housing. This report documents the first steps involved with inventorying the region's housing stock. The region's municipal assessors oversee the best-recorded information of housing information; that is, the grand list. Each year, the town's housing stock is recorded in the Municipal Grand List. This tax assessment database was linked with each municipality's GIS parcel coverage to provide a spatial location of the existing housing units, as well as information about the located housing units.

1.0 DIGITIZING THE 2000 GRAND LIST DATABASES

The 2000 municipal grand list data were joined with the most current parcel coverages for each town. Several methods were used to accomplish this. Methods varied from town to town due to variations among both grand list database structure and GIS parcel coding. As well, several GIS staff were responsible for this task.

1.1 Method One

For the towns of Bolton, Buel's Gore, Charlotte, Hinesburg, Huntington, Milton, Richmond, South Burlington, Shelburne, St. George, Underhill, Westford, Williston, and Winooski, the following method was used to join the 2000 grand list data to the GIS parcel coverage.

1. The grand list was copied into a DBASE format and named GLforMap.dbf
2. The following fields were added to the database: GL_KEY, GL_ONMAP, GL_RCRDCNT, and GL_RCRDNUM.
 - i. The GL_KEY is the value by which the record is to be related to the parcel map (the way this key is determined varies from town to town).
 - ii. The GL_ONMAP field contains a true value if the GL_KEY is actually on the parcel map.
 - iii. The GL_RCRDCNT field contains the count of grand list records that share the same GL_KEY.
 - iv. The GL_RCRDNUM field contains a unique value in the range 1 to the number indicated in the GL_RCRDCNT field. The GL_RCRDCNT and GL_RCRDNUM fields are needed because in most towns there can be multiple grand list records for a single map polygon. This is common in the case of condos or mobile home parks. For example, in Hinesburg the parcel 09-01-37.100 has 53 grand list records. Each such grand list record has a GL_RCRDCNT of 53 and a unique GLRCRDNUM in the range 1 - 53.
3. The town's parcel map was copied as a shapefile and named MapForGrandlist.shp.
4. The following fields were added to the polygon attribute table of the GIS parcel map: MP_KEY, MP_INGL, MP_RCRDCNT, and MP_RCRDNUM.
 - i. The MP_KEY field contains the value by which the record is to be related to the grand list (the way this key is determined varies from town to town).
 - ii. The MP_INGL field contains a true value if the MP_KEY is actually in the grand list.
 - iii. The MP_RCRDCNT field contains the count of map polygons that share the same MP_KEY.

- iv. The MP_RCRDNUM field contains a unique value in the range 1 to the number indicated in the MP_RCRDCNT field. The MP_RCRDCNT and MP_RCRDNUM fields are needed because in most towns a parcel can contain multiple map polygons. This is the case for example where a public road bisects a property. Note that for some parcels there are both multiple map polygons and multiple grand list records. This is the case for parcel 09-01-37.100 in Hinesburg that has 2 polygons and 53 grand list records.
5. Using MapInfo 5.5, these maps have been permanently joined using an outer join to the grand list in MapJoinedToGrandlist.shp. This file can be opened as a theme in ArcView. When the info tool is used in a polygon that has multiple grand list records, each such record is displayed. For example, the polygon attributed with MP_KEY = 09-01-37.100 in HINESBURG has 53 records associated with it. Clicking on this polygon in ArcView with the Identify Tool results in a display of 53 records. The following table illustrates the results of the outer join.

Map Joined With Grand List (Outer Join):

MP_KEY	MP_INGL	MP_RCRDCNT	MP_RCRDNUM	GL_KEY	GL_ONMAP	GL_RCRDCNT	GL_RCRDNUM
A	T	1	1	A	T	1	1
B	F	1	1	Null	Null	Null	Null
D	T	2	1	D	T	1	1
D	T	2	2	D	T	1	1
E	T	1	1	E	T	2	1
E	T	1	1	E	T	2	2
F	T	2	1	F	T	2	1
F	T	2	1	F	T	2	2
F	T	2	2	F	T	2	1
F	T	2	2	F	T	2	2

If a map polygon relates to N grand list records, there are actually N coincident polygons in mapjoinedtograndlist.shp. This means that mapjoinedtograndlist.shp **MUST NOT BE USED** for some purposes. For example, if one sums the areas of all the polygons in mapjoinedtograndlist.shp, the total area is wrong. Use MapForGrandlist.shp for such analysis.

1.2 Method Two

The 2000 municipal grand list data were joined with the most current parcel coverage for Burlington by the following process:

1. The most current parcel coverage was copied to a shapefile.
2. The MAPLOTNO, or map lot number, indicated in the parcel coverage was used to develop a geocoding address index.
3. Within the grand list, each record is given a unique identifier, the parcel ID. There is a separate parcel ID for each ownership associated with the parcel. The parcel ID is the map lot number plus a unique suffix.
4. The many-to-one link for parcels having more than one unit (and, hence, more than one owner), required the creation of two PARCEL_ID fields within the grand list:
 - i. The full individual PARCEL_ID. A unique PARCEL_ID exists for each owner.
 - ii. A truncated parcel ID. The abbreviated parcel ID is the same as the MAPLOTNO. Thus, the unique identifier for the owner was removed.

5. The truncated parcel ID from the grand list was then geocoded to the parcel coverage using the map lot number in the parcel GIS as the geocoded address field.
 - a. Using the truncated parcel ID from the grand list allowed multiple records associated with a single polygon to be joined with that polygon. At the same time, by retaining the full parcel ID from the grand list, the ownership and tax assessment information for each unit within a parcel polygon was maintained. Each individual grand list record can be identified via the full PARCEL_ID in the parcel attribute table.
6. The geocoded point locations for the grand list records were brought into ArcMap.
7. The geocoded points were joined with parcel polygon coverage using the Name/Attribute option. This was then exported as a polygon shapefile. Much as in the previous method, coincident polygons enable the many-to-one relationship between the grand list records and the parcel coverage. That is, if a map polygon relates to N grand list records, there are actually N coincident polygons in the polygon shapefile. Therefore, the polygon shapefile **MUST NOT BE USED** for some purposes. For example, if one sums the areas of all the polygons in the polygon shapefile, the total area is wrong.
8. The two PARCEL_ID fields have been maintained in the polygon GIS to allow for querying by individual PARCEL_IDS.

1.3 Method Three

For the remaining towns, Colchester, Essex, and Jericho, the method of geocoding was used. A GIS address database for the region was purchased from the vendor, DynaMAP USA. This is the same vendor assisting the Census Bureau in its geocoding efforts. Essentially, this is a map of the region's road network, broken into segments. For each segment, both the northbound and southbound (or eastbound and westbound) address number ranges occurring along that segment are recorded. Using a known address, a point is placed along the road segment that has an address range corresponding to the known address.

The P_LOCATION field in the grand list was used as the address field to match to the DynaMAP address. In Colchester, there was an 84% match of grand list records to addresses. In Essex, there was a 97% match. In Jericho, there was a 97% match.

1.4 Method Four

For all towns with E911 point data for structures, there was an attempt to geocode all un-located records to E911 points.

2.0 ACCURACY AND PRECISION

For all of the methods described above, there remained several grand list records in each town that did not correspond with the GIS parcel coverage. As well, there were several GIS parcels that did not correspond with any of the records in the grand list. These grand list records and GIS parcels were tagged for further review.

2.1 Percent Match

The table below lists the match rate between the grand list and the parcel coverage for each town.

Town	Total Records in Grand List	Total Records Matched to Parcels	Match Rate
Bolton	810	660	81.48%
Buel's Gore	31	27	87.09%
Burlington	9420	9152	97.15%
Charlotte	1826	1689	92.50%
Colchester	8059	7657	95.01%
Essex Junction	2611	2575	98.62%
Essex Town	3211	3186	99.22%
Hinesburg	1880	1763	93.78%
Huntington	953	895	93.91%
Jericho	2099	1497	71.32%
Milton	3705	3247	87.64%
Richmond	1649	1481	89.81%
Saint George	210	109	51.90%
Shelburne	2478	2429	98.02%
South Burlington	4785	3968	82.93%
Underhill	1481	1347	90.95%
Westford	927	561	60.52%
Williston	3088	2945	95.37%
Winooski	1627	1602	98.46%
OVERALL	50850	46790	92.02%

In each town, the unmatched parcels included roads, non-taxable parcels, surface water, railroads, and other municipal infrastructure. These were separated out from the other incomplete data and assigned a description, <NONE>, or <ROAD_ROW>, indicating the reason for their placement as “NA” (Not Applicable) in the dwelling type field of the dwelling units GIS attribute table. The other unmatched parcels were given “ND” (No Data) in the dwelling type field of the dwelling units GIS attribute table. In most cases, the records without data were later identified through work with the respective town office (see Section 2.3).

2.2 Unmatched Records / Missing Data

Each grand list-parcel GIS link was queried for the “ND” records in the dwelling unit type field. Maps and tables detailing the records for which data were missing were developed for each town. The means by which the missing data were obtained varied by town. The majority of the missing data were obtained using the municipal assessment database housed in the town’s computer, or by looking up the lister’s card for the parcel. Westford was particularly challenging because it was necessary to sit down with the lister and several maps to identify the parcels. CCRPC had old parcel IDs for Westford from before 911 road name changes. However, Westford has recently changed its road names for the purposes of E911, resulting in more than 1/3 of the parcel IDs needing to be changed.

In other instances, such as Colchester and South Burlington, digital orthophotos were used to identify the structure type and count on a parcel.

2.3 Random Sample

To ensure that the grand list matched with the parcels in an accurate manner, and that the linked information was correctly interpreted, the dwelling units database was verified using windshield and ground surveys. A random sample of the total database was generated and verified.

The dwelling units database is a special example of binomial data, in which the results of the parcel classification results in one of two outcomes, success or failure. The probability of success, or that the parcel was correctly identified as having a dwelling unit, is unknown but has an expected accuracy of 95%. To be certain that an adequate sample is used, it is assumed that the expected accuracy is 50% in the equation below. To determine an adequate sample size needed to verify that 95% of the parcels were correctly classified, the following formula was used (Senseman et al. 1995):

$$n = \frac{Z^2\alpha/2(p)(q)}{E^2}$$

Where n = total number of random samples needed, Z = 1.96 for the 95% two-sided confidence interval, p = expected percent accuracy, q = 100-p, and E = allowable error. Therefore, if we had an expected accuracy of 50% and an allowable error of 5%, the total number of points need is 384 (an expected accuracy of 50% was used to ensure that enough sample pints would be generated to allow for a substantial ground truthing effort throughout the entire county).

$$n = \frac{(1.96)^2(.50)(.50)}{(.05)^2} = \frac{(3.8416)(.50)(.50)}{(.0025)} = \frac{.9604}{(.0025)} = 384.16 = 384$$

Because each town has a separate grand list, there were several methods needed to obtain the link between the GIS parcel data and the grand lists. To ensure that each town is sufficiently sampled, the 384 sample points will be stratified by town. This would amount to approximately 22 points per town. Given the disparity in the number of parcels per town throughout the region, the sampling was weighted by the proportion of each town's total parcels relative to the total number of parcels in the region.

Of the 48,841 parcels in the region, there are 39,961 parcels in residential use. The following table shows the total number of parcels for each town, as well as the number of residential and non-residential parcels per town.

Town	Total Parcels	Residential Parcels	Non- Residential Parcels	Residential % of Total Parcels	Residential % of Regional Total of Residential Parcels	Weighted Sample Size	Sample Size - Rounded Up
Bolton	810	604	206	74.57	1.51%	5.80	6
Buels Gore	31	10	21	32.26	0.03%	0.10	1
Burlington	9420	8095	1325	85.93	20.20%	77.57	78
Charlotte	1826	1302	524	71.30	3.25%	12.48	13
Colchester	8059	5774	2285	71.64	14.45%	41.5	42
Essex	3211	2895	316	90.16	7.24%	27.80	28
Essex Junction	2611	2474	137	94.75	6.19%	23.77	24
Hinesburg	1880	1526	354	81.17	3.82%	14.66	15
Huntington	953	629	324	66.00	1.57%	6.03	7
Jericho	2099	1715	384	81.71	4.29%	16.48	17
Richmond	1649	1241	408	75.26	3.11%	11.93	12
Milton	3705	2873	832	77.54	7.19%	27.61	28
St. George	210	160	50	76.19	0.40%	1.54	2
Shelburne	2478	1990	488	80.31	4.98%	19.12	20
South Burlington	4785	3783	1002	79.06	9.47%	36.35	37
Underhill	1481	1146	335	77.38	2.87%	11.01	12
Westford	927	441	486	47.57	1.10%	4.24	5
Williston	3088	2222	866	71.96	5.56%	21.35	22
Winooski	1627	1459	168	89.67	3.65%	14.02	15
Totals	48841	39961	8880	81.82	100.00%	383.68	384

The overall accuracy of the random sample was 99.22%

The relative proportion of each town’s residential parcels to the total number of residential parcels in the region was calculated for each town. That proportion was used to weight the 384 sample points by town. The final column in the above table shows the results of this weighting.

An ArcView extension, the Randomizer, was used to employ the Ranshi algorithm. This was developed by F. Gutbrod, adapted from the Fermilab HEP Random library at (see <http://www.fnal.gov/docs/working-groups/fpcltf/Pkg/CLHEP/doc/html/CLHEP-random.html>). Tests to determine the randomness of the numbers generated with this algorithm were performed using the program DIEHARD (see <http://stat.fsu.edu/~geo/diehard.html>).

This script was loaded into ArcView and used to randomly select an appropriate number of parcels in each municipality. In some instances, the parcels cannot be fully seen from the road. Therefore, to avoid trespassing issues, those parcels were not checked and the next parcel in the randomly chosen list is used. For this reason, a number larger than the desired sample number of randomly chosen parcels were selected to ensure that the full sample for each town could be verified without imposing upon people's property.

2.4 Accuracy Assessment

With the exception of Colchester and South Burlington, a random sample of all sites was generated and then the sample was verified with site visits or a windshield survey. Several random sample generators were downloaded. Through a process of trial and error, the Avenue script MOA Random was used. This script was further refined and modified for each town, as the sample sizes and total parcel numbers are different. The sample size verified for each town was double the number indicated in the "Sample Size - Rounded Up" column of the chart in section 2.3. Each town's sample data were organized into a table that included the last name, street address, dwelling unit count, dwelling unit type, parcel acreage and sample number. In this way, the necessary information needed to assess the accuracy of the matched record when the actual parcel was located was available.

Colchester and South Burlington's final shapefiles still contain some records with a dwelling unit type of "ND", meaning "no data" and a dwelling unit count of "-99". These two towns were geocoded to make the match between the grand list data and the parcel coverage. The "ND"s and "-99"s indicate those parcels that did not make a match, and are therefore not complete. Colchester's sample was generated on only those records that were known to be residential. South Burlington's sample was done on all records. Colchester's final coverage is a point shapefile, while the others are polygons. Colchester did not have an adequate parcel coverage for use in this type of project.

The dwelling units shapefile attribute table for each town was augmented with fields for "Sample", "Found". The Y in "Sample" indicates that this record was part of the sampling process. "N" in this column indicates it was not part of the sample.

A "Y" in the "Found" field indicates that the parcel was located and the number of dwelling units present matched the data in the dwelling units attribute table. "NA" again means that the record was not sampled. An "N" in this field indicates that the parcel was either not found or the information in the attribute table differed from what was actually located on the parcel. If there were any "N"s in the "Found" column, a column called "Corrected" was added to indicate that the proper changes were made to the database. "Not Applicable" in this field corresponds to those records that were not part of the sample or those that were sampled but did not need correction.

The finished towns were converted to a final polygon shapefile. The parcel polygons were then converted to centroids using the ArcView extension, Xtools.

Particulars about some towns:

In Burlington, there were some inconsistencies in the mixed-use parcels around Church Street. Apartment locations were noted as having the address of the nearest ground-floor retail space, even when the apartment complex street number was visible at ground level. The records that were found to be disputed in this manner were noted in the “corrected” column. They were not corrected because it is not clear if the change is necessary. The dwelling units have been accounted for on the correct street, but not necessarily at the correct street number. In Colchester, parcels having multi-family structures with more than one record assigned to a single parcel will have records showing a dwelling unit count equal to 1 and a dwelling unit type of “MF” for multi-family. In order to find the number of dwelling units on the parcel, it is necessary to query the “Siteid” field (analogous to the parcel ID), which will show the same ID for records on that same parcel. These records can then be summed along the dwelling unit count field to obtain the total number of units located on that parcel.

Multifamily structures for all other towns have separate records for each unit except those in the following towns: Essex, Jericho, Milton, Shelburne, St. George and Williston, which have a separate table for some of the multi-family parcels. For these towns there were several instances in which an unmatched parcel was actually a condominium complex or a subdivision. The GIS parcel maps did not include the individual lots for the condominiums or the subdivisions. Rather, the GIS parcel coverage had one polygon to indicate the owner of the land on which the condominium complex or subdivision is located. It was not possible to alter the actual dwelling unit shapefile attribute table so that the number of records reflected the number of individual dwelling units. Rather, for these parcels, the dwelling units type is indicated as multi-family in the dwelling units database and the dwelling units count reflects the number of individual dwelling units in the complex or subdivision. For information on individual ownership, addresses, acreage and monetary values (where applicable), please see the corresponding table. All records that have information in such a table will indicate this in the “P_name_1” field of the dwelling units shapefile attribute table.

The following are the names of the multiple family information tables for the above towns.

TOWN	TABLE NAME
Essex	essexmf.dbf
Jericho	jerichomf.dbf
Milton	miltonsubcondo.dbf
Shelburne	terraces.dbf
St. George	stgeorgemhp1.dbf
Williston	willmf.dbf

Comparison with 2000 Census Data

Town	CENSUS 2000 HOUSING UNITS COUNT	DSS DWELLING UNIT COUNT	HOUSING DIFFERENCE (CENSUS-DSS)	ABSOLUTE HOUSING DIFFERENCE (CENSUS-DSS)	HOUSING % DIFFERENCE
BOLTON	412	433	-21	21	5.10
BUELS GORE		10			
BURLINGTON	15,885	18,785	-2,900	2,900	18.26
CHARLOTTE	1,287	1,358	-71	71	5.52
COLCHESTER	6,144	6,801	-657	657	10.69
ESSEX	3,604	4,059	-455	455	12.62
ESSEX JUNCTION	3,409	3,717	-308	308	9.03
HINESBURG	1,596	1,586	10	10	0.63
HUNTINGTON	692	710	-18	18	2.60
JERICO	1,751	1,793	-42	42	2.40
MILTON	3,333	3,503	-170	170	5.10
RICHMOND	1,504	1,432	72	72	4.79
SHELBURNE	2,632	2,257	375	375	14.25
SOUTH BURLINGTON	6,332	6,060	272	272	4.30
ST. GEORGE	264	271	-7	7	2.65
UNDERHILL	1,055	1,198	-143	143	13.55
WESTFORD	725	765	-40	40	5.52
WILLISTON	2,921	3,055	-134	134	4.59
WINOOSKI	2,944	2,723	221	221	7.51
COUNTY TOTAL	56,490	60,516	-4,026		(average) 7.17

Conclusion

This dwelling unit data was compiled using the best available parcel coverage and the 2000 grand list data for each town in Chittenden County. The information was checked for accuracy by windshield surveys of randomly sampled data. The accuracy of the final database is only as good as the source information and CCRPC is not responsible for any errors or omissions that may exist. This database was created with the intention of its use in the Regional Build-out Analysis and the Decision Support Systems projects. The map presentation of this data is not sufficient for delineation of features on the ground. It is in no way a replacement for engineering survey information. The map simply identifies the presence of housing units in relation to other features, and not necessarily their exact locations.