

**CT Broadband Mapping  
Data Processing Report  
Supplement**

---

**Submission 4**

**September 30, 2011**



## CONNECTICUT PROGRAM OVERVIEW

In response to the Notice of Funds Availability published in the Federal Register on July 8, 2009 (NOFA), the State of Connecticut Department of Public Utility Control (CT DPUC) submitted a grant application for consideration under the National Telecommunications and Information Administration's (NTIA) State Broadband Initiative Grant Program (SBI), for broadband mapping. The CT DPUC, pursuant to Executive Order 32-A, was designated as the single Connecticut state entity eligible to apply for funds under this program.

In July of 2011, the CT DPUC was merged with the CT Department of Environmental Protection to form a new agency called the Department of Energy and Environmental Protection (CT DEEP). CT DEEP will now be the lead agency coordinating with NTIA on this program.

The State has long been committed to regarding broadband delivery and enhanced use as a fundamental goal. The State has developed a planning strategy to marshal the State's resources and stakeholders and establish Connecticut as a leader in broadband usage, in addition to being a leader in "e-Government" and other broadband-dependent endeavors.

The State entered its SBI initiative not possessing any data related to broadband service, availability, or infrastructure that could readily support the requirements of the Broadband Data and Development grant program. Due to technical considerations, DEEP has partnered with Applied Geographics Inc., to support the data collection and mapping efforts.

So far CT has been very successful in acquiring the requested information from the broadband service providers, and is utilizing this information on our own <http://CT.gov/Broadband> website as well as providing the needed information up to NTIA to support the national map.

## FALL 2011 SUBMISSION OVERVIEW

For the fall 2011 submission (S4), roughly 75% of our providers submitted either entirely new or significantly revised data sets. This is a change from the last submission where half the providers just told us to reuse their existing data.

In general, the submission 4 processes followed the same basic approach that was used in earlier submissions. This document summarizes the following:

- Submission 4 Processing Assumptions
- Reference Data Creation
- Processing of new provider data
- Conversion from Census 2000 to Census 2010 format
- NTIA Submission Data Model Schema Changes

## SUBMISSION 4 PROCESSING ASSUMPTIONS

Based on NTIA feedback and information provided in NTIA webinar sessions, the submission 4 data processing workflow is based on the following assumptions to meet NTIA submission requirements.

1. All census blocks and road segments are mapped based on 2010 census data set. Any data submitted in 2000 or 2009 format was converted to 2010 for submission.
2. Overview was removed completely from this submission data due to the fact that all maximum advertised up/down speeds are being reported in blocks, roads, and wireless features. In addition, none of the providers were willing to submit detailed pricing information.
3. Due to our NDA restrictions, last mile points will not be submitted to NTIA.
4. Due to NDA restrictions and our inability to accurately flag service by “category of end user”, address points were not submitted to NTIA for any commercial provider.
5. Some providers did not submit middle mile elevation. Wherever possible, we went back to providers to obtain their middle mile elevation information, but it is not available for every record.
6. Terrestrial Mobile Wireless and Terrestrial Fixed Wireless (licensed and unlicensed) were treated as wireless coverage and were delivered as a shape. In cases where a provider served the same technology and spectrum with different speeds, overlapping areas were removed and the higher speed was assigned.
7. If a cable based wireline provider can provide both DOCIS 2.0 and DOCIS 3.0 service to the same area, the block or road was listed only once with a technology code of 40.
8. Providers were only willing to indicate on a general level if they served business, residential or both, so we did not get any providers that broke down the type of service by block. Only if the provider stated they only serve business to business customers did we fill in the “category of end user” with a code of 2, otherwise this field was left blank.
9. The submission 4 Provider data model is currently based on the NTIA data model as of 6/30/11.

## SUBMISSION 4: REFERENCE DATA

This section describes the reference data used in submission 4.

### BLOCK REFERENCE SETUP

For s4, Census 2010 data was utilized. The data was set up as follows:

- Block size (AREA) is calculated combining the 2000 land area (ALAND) and water area (AWATER)
- AREA is converted from square meters to square miles to calculate square mileage (SMI).
- If the SMI of a block is less than or equal to 2, then the less than or equal to 2 square mile indicator (LE2SMI) is set to true.

### ROAD REFERENCE SETUP

2010 Tiger Line IDs (TLID) were used for data processing in s4. The data was set up as follows:

- The GT2SMI (Greater Than 2 Square Mile) indicator is set to True when:
  - The 2010 road segment is completely within a block that is NOT less than 2 square miles
- Only minimum and maximum address ranges and a single zip code for each road segment is maintained.

## SUBMISSION 4: PROCESSING OF NEW DATA

For submission 4, AppGeo started data collection on July 6<sup>th</sup> 2011 by sending out data update requests and technical data specifications to all providers. This incorporated all the NTIA changes released on June 30<sup>th</sup>, 2011. These were sent to a large list of companies which were compiled from past collection efforts, and the revised FCC 477 list. The technical document highlighted the changes from Submission 3 to Submission 4. Due to the change in census geography all new data was requested whenever possible.

We then actively followed up with the providers. As we had discovered in the past, many of the providers listed on the FCC 477 list are either resellers, or not involved in the actual delivery of broadband. (Many are VOIP or teleconference service providers that utilize existing broadband connections.)

In our solicitation for data updates, we told known past providers that if we didn't hear from them by a certain date, we would default to using their data from Submission 3. We contacted them after the due date a few times but for two providers, we eventually had to just reuse Submission 3 data.

All data received went through the following processing steps:

1. **Triage:** All new data was quickly reviewed to understand what was received, and in what format. We also made sure we had all the required components for NTIA's data model, such as their FRN and advertised speed information. We also screened for any known issues that we might have seen before (such as Excel 2003 spreadsheets that cut off at 32k rows.)
2. **Ingest:** At this time the data is actually brought into our systems. Each provider is set up with a unique file geodatabase to store their information. Record counts of what was received is logged so that we can validate we did not drop anything in processing.
3. **Data Processing:** This is where the data goes through a number of ETL routines to convert the raw proprietary information into a format similar to the NTIA format. The exact routine utilized depends on how the data is received:

- a. When a wireline provider submits a service boundary, we select all the blocks and roads inside that shape.
  - b. If a wireline provider submits a customer address list, the points are geocoded, and then the appropriate block or road segment is selected.
  - c. If a wireline provider submits block and road information using Census data, we just make sure everything is formatted to the appropriate specifications
  - d. If the wireline provider submits any type of road or line data that does not directly correlate to the TIGER data set, we convert the lines to TIGER by selecting the road centroid and spatially selecting the closed segment in our data set. If the road is in a block less than 2sqmi, then the block is selected. Some manual cleanup is also applied to make sure we do not accidentally drop any road segments that should have been processed.
  - e. Wireless provider data is formatted to ensure that there are no any overlapping polygons with the technology type. In addition the data is cropped to the state boundary.
  - f. After each round of processing, we make sure that we only keep unique records. A unique record is defined as having a one of a kind combination of FRN, Block/Road ID, and technology type. If there are multiple records with different speeds, but all else is equal, then we select the maximum of the advertised speeds.
4. **QC Review:** All data is then sent to a different analyst to perform a thorough quality control review on the processed data set. Record counts are compared to what was submitted. The QC staff also make sure the ETL scripts and routines populated all of the right fields.
5. **QA Review:** Data is then sent to another team for Quality Assurance Review. In this step the data is not only double checked against what was originally submitted, but it also brought up inside standardized MXD templates that allow us to make sure our results make sense. This often involves comparing the new data set with prior submissions, as well as looking for any possible technology or speed anomalies. At this stage we also start in on our validation process. This includes looking at the provider data in comparison to things such as speed test results, franchise boundaries, siting information, and feedback from the planning surveys.
6. **Provider Review:** Processed data is all posted to a customized web application we refer to as our Provider Portal. All providers were notified once their data was available in the site, and they were always given at least ten business days to review the data and respond. In this site, providers can log on and visually see their processed data in a map format. It also allows them to overlay their raw data to help them validate that we did indeed process things correctly. The provider portal also has a suite of markup tools that will allow the providers to edit their data, including adding or removing service areas, and making changes to the data attributes.
7. **Comment Processing:** All comments and feedback received from the provider portal, is then reviewed and applied to the processed data set. This updated data set goes back through our QA and QC processes, and if time allows, back out to the Provider Portal, for the provider to review and sign off on.
8. **Data Append:** After all of the individual data sets are processed and approved, we run an append process which merges all of the individual provider data sets into one geodatabase. This is also the point where our team will do any final transformations to get our working data model into the latest NTIA publishing format.
9. **Final QA/QC:** A series of quality checks are run on the final appended data sets to ensure it is ready for submission to NTIA. We also run the NTIA receipt tool at this time. Any last issues are corrected, and the data is sent to the state for their review.
10. **Submission to NTIA**

As with the third data submission, we followed the following protocols:

1. We did not collect data from resellers
2. We collected data from satellite providers, only if they were able to provide to us all of the required information we need to pass onto NTIA: including spectrum, FRN, and advertised speeds.

## COMMUNITY ANCHOR INSTITUTIONS DATA

The community anchor institutions data was primarily populated through State resources, in particular the CEN database which services many schools, colleges, and libraries. We also were able to get a connection survey results for all the libraries through the state library association. Location information for all other CAI points, notably, police, fire, and town halls, were obtained through the Department of Public Safety. All of this information was then populated into an online data gathering and validation web based application. Each town was contacted and asked to update their respective site information. While the web based responses have not been as high as we would like, we do feel that we are fortunate to have a good base set of data from the state.

## SUBMISSION 4: 2000 TO 2010 DATA CONVERSION PROCESS

Due to the changes in census geography, all providers were asked to submit new data. In those instances when a provider A) submitted new data in Census 2000 format, or B) instructed us to reuse their last data submission, we had to convert the blocks and roads into 2010 format.

### Basic 2000 to 2010 Conversion Process:

1. For the blocks, take the 2000 block ID, and select all the corresponding 2010 block id's
  - a. using census crosswalk table – not an actual spatial process, since this was faster
2. Look at the new 2010 block ids, and filter on greater than or less than 2 sq miles.
  - a. If less than or equal to 2 --> bring in the 2010 geometry and add that record to the blocks table
  - b. If greater than 2 --> select any roads in that area – spatial select (using roads gt2 table)
3. For the roads, take the 2000 or 2009 TLID and try to match it to the 2010 TLID's
  - a. If there is a match, add that record to the roads table
  - b. If there is not a match, select centroid of existing 2000/2009 segment, and select closest 2010 road
  - c. If the road is now in a block LT2, select the block(s) instead and drop the road
4. Remove any duplicate records in both tables
5. Run some automated checks to catch missed features (i.e. add le2smi blocks surrounded by roads that have not already been added)
6. Manual review (QC) and corrections.
  - a. There will be some blocks that are selected inappropriately (especially at town edges for CT providers, where we know their franchise ends at a town line.)
  - b. There are some holes in the census crosswalk table
  - c. The road conversion process may only select one portion of the road if it has now been broken into multiple segments

### Assumptions

1. If a road was in an area greater than 2smi in s3, and due to census re-drawing, is now in an area less than 2smi, we will grab blocks (le2smi) on both sides of that road and add them to the provider data:
2. If a new 2010 block, that is less than 2smi, is completely surrounded by roads and/or blocks served by that provider, than we will add the block to the provider service area.

## SUBMISSION 4: NTIA DATA MODEL SCHEMA CHANGES

The data model released on June 30, 2011 contained the following changes to the s3 data model:

- The Category of End user field was added back in to the block and road tables. In addition the domain values were changed. 1 still represents residential, but a 2 now represents all non-residential uses.
  - This field is not required, and for many providers, was left blank since the data was not provided.

## CONNECTICUT SPECIFIC INFORMATION

Due to Connecticut's geography and population, 99.75% of the census blocks in the state are less than two square miles. The need for us to break apart coverage based on blocks versus roads leads to a lot of unnecessary confusion as well as creates some distorted pictures when you try to visualize this information on a map. For this reason, all of the maps available on the [CT.gov/broadband](http://CT.gov/broadband) website are published after we convert all of the data to just use blocks.

In the documentation form NTIA there has been a lot of discussion about making sure that a provider uses the same DBA and FRN consistently across all feature classes. We mentioned this to the providers, but there was some push back. Most providers complied with this request, but a few providers pointed out that while they may share a common name, they actually operate as separate organizations. Also, due to regulatory implications of the different FRN's a few providers did insist that their records not be combined.

The State of Connecticut has built and maintains the Connecticut Educational Network (CEN) which is a mostly fiber backbone connecting educational facilities all across the state to the internet. The way this network is set up, it is difficult to accurately identify who the final provider is and at what speeds. In particular, the CEN network will typically install one fiber uplink in each town, and then it is the town's responsibility to provide connection between facilities. So for example CEN may supply the board of education's office with a 10mb connection, but then the board of education will run lines to each of the schools in the district. So towns are reluctant to report speed information as there may technically be 10mb available to the school, but reporting that speed at each school would grossly overestimate how much connectivity they have in total, when in fact there may be 15 schools sharing that same uplink.

Comcast cable has reported that they market their speed offerings based on MSA areas, which do not necessarily match up with their technology available in the field. So what this means is that there are some areas that may in fact have DOCSIS 3.0 technology available, but the market they are in is still offering packages and speeds based on the older technology. This means you will see areas with technology type of 40, and advertised speed tiers of 7 or 8. We did work with the provider to make sure that any area that has technology type of 41 is not using the DOCSIS 3.0 speed offerings.

One Communications Corporation has a few locations where they offer what they called stacked DSL service, and they confirmed in writing that these locations can offer aDSL with speeds in tiers 7 or 8.

We have already noted that PAETEC has listed most of their locations with an advertised speed of tier 10, but in many cases are reporting typical speeds in tier 2. We have contacted the provider and they have simply responded that we should use the data as it was submitted.