

“White Paper” from *New York* describing Round 4 (Fall, 2011) Data Submission to the NTIA under the SBDD

October 1, 2011

Executive Summary

The Broadband Mapping Team at the New York State Office of Cyber Security (OCS) is pleased to submit our Round 4 (Fall 2011) data for the State Broadband Initiative (SBI).

Our goals for Round 4 were to: 1) maintain the very high level of participation from New York providers, 2) Add to and enhance our data verification methods and, 3) improve the completeness and quality of the data delivered. We believe we have met those goals.

We had 82 providers participate in the spring 2011, Round 3 submission. That number has risen to 87. We anticipate an increase in that number in the future as we continue to reach out to small fixed wireless companies. We believe mapping these provider’s serviceable areas is a very important component required to fine tune NYS’s served and un-served boundaries.

We are very pleased with the enhancements to our verification methods implemented this round and are energized to continue to find innovative ways to use disparate data available from commercial, government and public sources to validate and identify inconsistencies in provider reported availability data.

Lastly, we made small but significant steps in improving the Community Anchor Institution data (attributes and location), middle mile points and, by working even closer with our providers, we were able to improve the quality of the availability data using the new 2010 Census geography.

The remainder of this paper provides a summary of our data collection results; describes our methodology for performing data verification; summarizes the progress made on all our Round 4 goals and identifies our focus for Round 5.

Provider Participation Summary Tables:

| | |
|-----------|---|
| 87 | Total Participating Providers |
| 72 | Wireline Providers |
| 17 | Wireless Providers (2 are both Wireless & Wireline) |
| 1 | Provider is middle-mile only |
| 44 | Providers submitted Middle Mile Data |

| Technology Type | Wireline Census Block Provider Count | Wireline Service Availability by Census Block | Wireline Street Segment Provider Count | Service Availability by Street Segment | Wireless Provider Count | Wireless Services by Shapefile | Middle Mile Provider Count | Middle Mile Points |
|---|--------------------------------------|---|--|--|-------------------------|--------------------------------|----------------------------|--------------------|
| Asymmetric xDSL | 43 | 307,636 | 35 | 26,945 | 0 | 0 | 28 | 1,704 |
| Symmetric xDSL | 6 | 66,616 | 2 | 81 | 0 | 0 | 0 | 0 |
| Other Copper Wireline | 7 | 93,947 | 4 | 235 | 0 | 0 | 1 | 4 |
| Cable Modem - DOCSIS 3.0 | 8 | 197,110 | 6 | 15,069 | 0 | 0 | 3 | 10 |
| Cable Modem Other | 13 | 173,283 | 11 | 21,022 | 0 | 0 | 1 | 1 |
| Optical Carrier/Fiber to the End User | 23 | 120,961 | 14 | 2,129 | 0 | 0 | 7 | 652 |
| Satellite | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 |
| Terrestrial Fixed Wireless - Unlicensed | 0 | 0 | 0 | 0 | 7 | 12 | 1 | 10 |
| Terrestrial Fixed Wireless - Licensed | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| Terrestrial Mobile Wireless | 0 | 0 | 0 | 0 | 6 | 13 | 2 | 14 |
| Other (middle-mile only) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |

A note regarding New York's Provider Table in Data Package XLS:

In the July 13, 2011 "NY December 2010 Data Feedback" conference call, the program office requested that New York include in its submission any company that we had researched in order to determine if they met the "provider" definition of the program and provide services in NYS. For this submission, we have included an exhaustive list of providers that we have captured information on from Rounds 1 through 4.

In our Provider Table, companies are listed by order of participation status with those that "Provided Data" listed first, followed by companies that said they "Will Provide Data", those that were "Non-Responsive", and finally those that "Will Not Provide Data." Providers that were contacted in Round 4 are contained within the top 3 statuses and include all known facilities based providers.

Companies that received a "Will Not Provide Data" status include the following:

| Company Type | Code |
|---|------|
| Providers that do not serve New York | 3 |
| Companies who are not broadband providers | 4 |
| Broadband equipment companies | 3 |
| Resellers | 2 |

Because of the sheer number of resellers in New York and their questionable ability to provide service within 7 to 10 days, resellers were a low priority for Round 4 outreach.

The companies that populate the Provider Table were gleaned from the FCC list provided in the "NY SBDD Submission Summary.xls" as well as from our own research in Rounds 1 through 4. We hope that this list is useful for the program office and provides some clarity on the volume of providers New York has researched.

Verification:

Automated verification was accomplished via the following methods:

1. Business rules built into the SBI data transfer model (catching problems on the way in)
2. Repeatedly running the NTIA supplied Python script
3. ESRI 'Check Geometry' and 'Fix Geometry' tools

Non-automated verification methods ranged from the very simple to complex, multi-step procedures. They were:

1. **Provider Website Research:** For Round 4, a careful study was made of the websites of New York State's 80+ broadband providers. The goal was verify the technology and data transfer speeds that were voluntarily self-reported to the OCS broadband mapping team.

Upon this review, a few generalizations can be posited and some interesting facts can be gleaned. First, it has to be noted that although there are at least 83 separate companies offering DSL, cable, fiber, or wireless technologies, there were actually only 67 differing websites to examine. This is due to the fact that some of the smaller companies have the same parent company. Thus, there is one uniform website for all of the parent company's subsidiaries. For example, when searching for the main website of Berkshire Cable Corporation, Berkshire Telephone Corporation, Chautauqua and Erie Telephone Company, or Taconic Telephone Company, the same exact FairPoint Communications website will pop up. If there is some disparity in the products and speeds per individual company, it cannot be discerned from the catch-all corporate website.

This study exposed some noteworthy details. 12 of the companies provide no specific information on the actual data transfer speeds of their broadband products. Two of them, Castle Communications of Willsboro, and Fishers Island Telephone Corporation, did not even have a working website! Of the remaining 71 companies that did provide speeds, only 29 of them provided both downstream and upstream speeds.

Once all the website speed data was collected from the websites, the maximum download and upload speeds offered were reclassified and coded into speed tiers between 1 and 11. The vast majority corresponded to self-reported speeds; however, there were some notable discrepancies.

Most discrepancies were within one or two code numbers. For example, Verizon reported to us that they offered "6" speed broadband at Tech Code 10 (ADSL), and their website indicated it was "7" level speed. There were a handful of companies though that reported much higher speeds than what was found on the internet. Often, they were companies that offered broadband at a Tech Code of 50, which indicated optical carriers/fiber to the end user. Thus, they offered high-end technology or specialized technology producing higher speeds that was not detailed on their site. For example, DFT Local Service Corp reported to us they have "11" speed technology, but it is not detailed on their website, as we have uncovered only code "7" level speed, which is what their fastest DSL service registers.

There were approximately 26 cases where speed code differences ranging from minor to significant occurred. In half the cases, the max speed advertised on the company's internet site was less than what they reported to us. Therefore, the other half of the cases involved the internet sites boasting

higher speeds than what we were aware of, so it is difficult to ascertain a pattern. In 8 of the 26 instances, there were speed code disparities of 3 or more. For example, Hometown Online, Inc. reported to us maximum speeds in the 8 speed tier range, but their website indicates speeds up to 2 mbps, which is only within the 4 speed tier. It should be noted that this data is just comparing maximum download speeds, because so few maximum upload speeds were advertised on the websites.

While the results on this study were not conclusive enough to permit significant revisions to the providers' self-report speed data, this study will serve to generate constructive discussion with our providers and assist with future refinements to our speed data to more accurately portray the diverse palette of broadband coverage offerings.

2. Use of crowd-sourced data:

- a. NYS Speed Test data points and attributes were used to verify provider reported availability.

The NYS speed test website includes a data collection form which requests:

- i. Street address at which the test was taken
- ii. Service provider
- iii. Service technology

After satellite provider records and sub-broadband speed records were removed, 5624 records were successfully geocoded and used for verification. Four levels of verification were established for points that fell within areas of reported service availability. They are:

Code 1 = Provider and technology matched

Code 2 = Provider matched and technology unknown

Code 3 = Provider matched but technology is mismatched

Code 5 = Provider and technology unknown but Broadband is available in the location

Each census block and street segment availability record involved with this verification activity was assigned one of the above codes.

- b. FCC speed test records were used to verify provider reported availability. FCC speed test records lack provider information but we were able to successfully establish the provider via a publically available IP Address search engine (the APNIC Whois Database). Those records were then used to verify provider reported availability in the same manner as was used with the NYS speed test points. Because the technology was not known, the highest verification code assigned was 2 (Provider Matched and Technology = 'Unknown'). Here is a statistical summary:

| | Number | Percentage |
|--|-----------------|------------|
| Total Number of FCC Wireline Speed Test Points | 62,642 | N/A |
| Total Number / Percentage Successfully Geo-coded | 32,621 / 62,642 | 52% |
| Total Number / Percentage Successfully IP Searched | 21,766 / 32,621 | 67% |

- c. NYS Broadband Map feedback: After receiving an email through the “Is This Correct” link on the NYS broadband map, the details were logged in a tracking spreadsheet and investigated on our map. The address, census block, or street segment was then further investigated in ArcMap using provider submitted data to confirm reported availability. If we confirmed that the provider submitted availability data for that location, the next step was to use the provider’s own website to attempt to verify that availability.

If available, the public responder’s address was used along with address point datasets from New York State and Navteq. In a census block or street segment, addresses were identified at both ends of the bounding features. These addresses were entered into an availability search on the provider’s website and the results were logged. In Frontier’s case, the address points were used to perform a reverse lookup and identify phone numbers at those addresses. The phone number was then entered on Frontier’s site.

If an address within the block or segment was identified by the provider’s site as potentially served, that block or segment retained that provider’s coverage on our map. If no addresses within the block or segment were identified as potentially served, we removed coverage of that block or street segment for that provider from our map.

This round, for the first time, we investigated surrounding “suspect” blocks and segments. These are areas that were submitted as served by providers, but contradicted some of the “on the ground” knowledge we received from the public. These were also areas that stood out spatially (i.e. non-contiguous or “island” coverage) against the type of technology in question (i.e. wireline technologies that run along roadway).

Here are summary statistics for this feedback activity:

| | | Number |
|---|---------------|----------------|
| Public emails received during Round 4 | | 102 |
| | Block Records | Street Records |
| Number of locations investigated and verified | 87 | 4 |
| Number of locations investigated and removed | 130 | 212 |

3. **Use of newly identified government data sources**: The NYS Department of Motor Vehicles supplied three new datasets for our verification activities. Satellite Offices, Dealer Locations and Inspection Station Locations were used to verify provider reported availability. *All of these facilities have broadband connections.* The Dealer and Inspection Location datasets did not have provider or technology information associated with the locations. Therefore, the highest verification code assigned was a 5 because we were only able to confirm that there was broadband at those locations. However, the DMV Satellite Offices dataset came with provider information, so those locations were assigned a verification code of 2.

4. **Use of newly identified commercial data source:** TomTom data was used to verify provider reported availability. The TomTom data included boundaries for many of the broadband providers we have received data from. The TomTom boundary for each provider included in the dataset was overlaid onto the provider footprint from our SBI data transfer model. This was done to ensure that the availability data sent to us by the providers was within the respective boundary in the commercially available TomTom data. All of the provider footprints that had matching boundaries in the TomTom data fell within their respective boundary.
5. **Select CAI locations** were used to verify provider reported availability for the first time this round. We selected Colleges, Hospitals, Federal Correctional Facilities, State Prisons and State Police Stations from our total collection of previously identified CAIs as an additional verification data source because we strongly believe all have broadband connections. Since the provider and technology are still unknown at this time, the highest verification code assigned was 5 (Provider and technology unknown but Broadband is available in the location). All but 12 of the 887 points used confirmed provider reported availability. The 12 conflicts are currently being researched. Results will be reflected in Round 5 data.
6. **Provider verification:** For providers with significant changes from the previous round, we created review maps showing Round 4 availability aggregated to census blocks and street segments. There providers were given at least five days to respond and initiate any changes or corrections. Changes were made based on provider feedback. Changes were documented for future reference. These OCS generated maps were later compared to the provider footprints in the geodatabase to ensure that the data loaded in correctly. Many of the providers have multiple review maps, so each of these maps had to be examined and compared to the corresponding area in the data. During the comparison process, four provider's footprints were discovered to have some missing data and were corrected. The rest of the provider's footprints matched their respective review maps.
7. **Verizon NY (wire-line) specific scrubbing:** Verizon New York submitted data in 2010 TIGER/Line Census Blocks and Edge Files as text delimited files.
 - a. Street segments in the original data were highly fragmented and discontinuous in census blocks greater than 2 square miles. An infill process was used to select segments 100 meters or less where availability was not reported by Verizon but that segment fell in between two street segments with Verizon reported availability. Addresses from a sampling of the new street segments were checked through Verizon's website and broadband availability was verified. A total of 920 segments were added to Verizon's availability and assigned the max advertised speed attributes of the nearest street segment. 312 Street segments that were discontinuous with any other reported availability or fell more than one mile outside Verizon's service footprint were checked for availability through Verizon's website and subsequently deleted.
 - b. Census blocks in the original data contained outliers. 122 Census blocks reported by Verizon fell more than one mile outside their exchange boundary and were discontinuous with any other Verizon reported availability. Addresses sampled in these blocks were checked for availability

through Verizon's website. Through this process, it was verified that there is no Verizon service in these blocks, and the census blocks were deleted.

8. Clipping all data to the NYS boundary file

Round 4 Anticipated Focus and Current Status:

The six items below describe work we believed, at the time of our last submission, warranted special emphasis during the next round of data collection, processing and verification. The current status is provided after each description.

1. *From April 1, 2011 methodology paper:* Further attribution and enhanced spatial accuracy of our Community Anchor Institution (CAI) data: To date, collection of the broadband service attributes for our CAIs remains one of our activities in need of the most improvement. To that end, we are nearing a final version of a proposed scope of work for our partner, the Center for Technology in Government (CTG), to expand their data collection activities beyond speed test data to include CAI broadband service attributes. The speed test data collection website and CTG's outreach network will be further leveraged.

Current Status: We amended our Memorandum of Understanding with CTG to expand their outreach and data collection activities to include broadband service attributes for schools, libraries, colleges/universities, medical/healthcare facilities and municipal halls (Other community support – government category). CTG successfully used their outreach network to bring the State Education Department (which governs schools, colleges/universities and libraries) to the table and agree to begin a joint broadband data collection effort once per year. This will consolidate three separate survey efforts that collected basically the same information and should result in a very high rate of participation within these sectors of CAIs. This first joint collection is scheduled for October 2011. OCS was provided earlier survey results from these sectors and we were able to extract partial SBI required attributes for use in this submission. CTG will also focus on municipal halls and medical/healthcare institutions for the next submission, OCS will focus on the public safety sector. While it took a great deal of time to get this far, this is a significant step in the right direction for this area of focus with collateral benefits for the State Education Department as well as the individual institutions that must complete state surveys.

2. *From April 1, 2011 methodology paper:* Identifying and working with fixed wireless providers: We believe we have yet to identify some existing providers and new companies will be starting up to fill small pockets of underserved or un-served.

Current Status: In Round 4 we reached out to 8 companies we identified as fixed wireless providers in New York. In most cases these companies were non-responsive. Two companies, Logical Net Corp. and NY Air, expressed willingness to participate. However, after our initial exchanges, both companies

became non-responsive. We had a similar occurrence in Round 3 with the fixed wireless company Plexicomm. We believe that the small sizes of these companies, as well as the perceived amount of effort to provide data, are the greatest hurdles. New York plans to keep reaching out to these providers as they are seen as a viable alternative to wired broadband in underserved areas and areas previously thought to be unserved.

3. *From April 1, 2011 methodology paper:* Adding verification methods: We intend to pursue the use of additional crowd sourced, commercial, and public data source and the aggregated FCC supplied 477 data.

Current Status: As noted above, the use of the crowd sourced data received via our state broadband map was extended to verify “surrounding suspect areas”. This resulted in additional corrections that would have otherwise not been realized. No new crowd sourced data sets were identified or created this round. We also utilized a newly discovered commercially available TomTom data product to further verify provider footprints DMV (see details in verification section above).

We researched InfoGroup (formerly InfoUSA) data. One cut of data they offered contained “marketing” records with a single broadband related attribute (email addresses that in some cases could be associated with technology). They also offered to query their data store using NAICS codes in attempt to identify CAIs and provide contact info, location address and email address. After reviewing sample data, we decided neither offering was going to be very helpful.

We obtained and utilized three new datasets obtained from the NYS DMV (see details in verification section above). Most of this information is publically available but the DMV provided it to us in a “ready to use” format. Lastly, we are nearing the end to the process of agreeing to and satisfying the conditions for obtaining the aggregated FCC 477 data. We will be able to utilize this data for verification of Round 5 data. We will develop methods using Round 4 data.

Overall, use of additional data sets and extending our verification methods have allowed us to get closer to our goal of eventually having record level verification of all data within the SBI data transfer model deliverable. To ever reach that level though, we will need to automate at least some of our currently manual methods but we continue to get new, innovative ideas each round.

4. *From April 1, 2011 methodology paper:* Migration to 2010 Census data layers: This will involve the realignment of new Census geography to NYS basemap layers and migrating the previous round’s data to Census blocks that have entirely new id numbers.

Current Status: A migration to US Census Dept. provided 2010 geography was completed. All block level data delivered in the SBI data transfer model is attached to 2010 census blocks. Street level data is attached to the most current version of the NYS streets geography. OCS will likely contract out the work to have the Census supplied geography aligned to NYS streets and other NYS base map layers. We receive an estimate from Navteq to perform this realignment and conflation work in time to use these

results for this Round 4 deliverable. Navteq is currently under contract with OCS and there was the potential to leverage that existing procurement vehicle. We felt the cost was too high, in large part because of the “rush” nature of our request. We declined and will likely have to do a competitive procurement to obtain an “affordable” price.

5. *From April 1, 2011 methodology paper:* Improvements to the NYS Broadband Map and increasing the number of ‘visits’: We see our state map as an area where we can provide value to our provider partners. We have already met with some providers to discuss displaying multiple ‘speed package’ offerings. Time Warner Cable has agreed to work with us to pilot that enhancement. We are also in discussions with CTG in order to have them perform outreach work to increase the visits to the site and specifically encourage visitors to provide feedback regarding the accuracy of the availability data. We already have a detailed verification workflow in place to effectively utilize this data.

Current Status: We made changes that we believe enhance the user-friendliness of our State Broadband Map, including a homepage redesign (www.broadbandmap.ny.gov). We also added provider footprints to the map. With our October 2011 release of the Round 4 data, we will be adding functionality to view multiple speed offerings data voluntarily given to us by 11 providers (including Time Warner, a major NYS provider). We will also be adding a composite max advertised speed layer and discrete single speed tier layers. Functionality allowing the public to identify unserved addresses is planned for release before the end of this year. CTG has agreed to expand their scope of work to include marketing our mapping site and feedback tools. A second amendment to our MOU with CTG is nearing completion.

6. *From April 1, 2011 methodology paper:* Further development of a project plan for our address point development work: We are already using address points for geocoding service delivery addresses and for verification work. For Round 4, we envision our use of address points for verification to increase and for their use in enhancing our ability to estimate household availability, underserved areas and uninhabited lands. Needs assessment discussions are already underway with E911 and key government agency stakeholders.

Current Status: The Address File portion of our project plan was expanded and approved along with our entire project plan, budget and timeline package required under our supplement grant award. We have recruited two county 9-1-1 organizations to work with us on a pilot to develop address point mapping standards and define data workflows. Work on that pilot is expected to begin before year’s end. We also completed work with a professional demographer reviewing our methods for calculating household units with broadband availability. A final report was produced and posted to PBWORKS.

Round 5 Focus

1. One of our primary focuses will be to develop methods to reduce the upward bias in our calculation of household units with broadband availability. We will continue to work with the demographer we have under contract as well as further explore our own ideas. When NY availability percentages are

presented at public forums, we consistently get feedback that our numbers are inaccurate. At the most recent event attended, Congressmen Gibson's Rural Broadband Symposium (9/29/11, NY's 20th District), US Department of Agriculture staff were also presenters. When discussing their Rural Development Community Connect grant program, they stated that in 100% of over 40 instances where an applicant's claim of being unserved was in conflict with the data on the NBM, the applicant's claim was substantiated by Dept. of Ag. local field staff investigations. Clearly the granularity of data collection is insufficient to produce accurate mapping in rural areas of the country. We see improvement here as paramount to the continued success of our state's program. We anticipate increased use of public feedback and other crowd sourced data playing a significant role in improving the accuracy of the availability data.

2. CAI attribution: Collaboration agreements with multiple units within the NYS State Education Department and a refinement of our CAI definition have laid the ground work for significant advancement in this area.
3. Advancement of our Address Point Mapping Standards pilot and additional use of address point data for data verification: Our work to date with the NY 9-1-1 Coordinators Association and a related pilot project currently being planned with US Census Bureau has set the stage for significant advancement of our Address File project.
4. Additional enhancement of our verification activities: This will include the addition of "unserved address" functionality to our state broadband map; the use of the aggregated 477 data and the use of yet to be identified new data sources.