"White Paper" from *New York* describing Round 3 (Spring, 2011) Data Submission to the NTIA under the SBDD

April 1, 2011

Executive Summary

The Broadband Mapping Team at the New York State Office of Cyber Security is pleased to submit our Round 3 data for the SBDD Program. Our goals for Round 3 were to: 1) obtain the required data elements from all known facilities-based providers of broadband service to end users in the State of New York, and 2) improve the quality of the delivered data through greater emphasis on geospatial verification procedures and the use of additional sources of information in the data verification processes. We believe we have met those goals.

Going into this round of data collection, we had concerns that providers' interest may have waned and that it could be difficult to retain the current participants. Through hard work and continued emphasis on building partnerships, we retained all but one previous participant and gained six new participants. Our Round 3 data includes data from 100% of known facilities-based (wireline) end user providers and all significant wireless service providers in NYS. We have one small fixed wireless provider who participated in the previous round of data collection who indicated that they did not have the resources to continue with the program. We will seek a solution to bring this provider back in for future rounds.

Overall, we are extremely pleased with the net gain for Round 3 participation!

Although we did not target broadband resellers for Round 3, our delivery includes data from six providers who do not own their own facilities. We anticipate expanding the range of program participants to include more of these resellers in future rounds of data collection. Guidance from the NTIA program office and information sharing from within the SBDD Grantee community will be helpful in guiding these future data collection and mapping efforts.

Another future goal is to identify and work with more fixed wireless providers. Our experience thus far has shown fixed wireless providers to be mostly small, relatively new companies targeting geographic areas where un-served pockets left by wire-line or cellular wireless companies exist. We believe mapping these provider's serviceable areas is a very important component required to fine tune NYS's served and un-served boundaries.

Expanding our data verification methods was another area of focus during this round. Just prior to our Round 2 data delivery, we launched a speed test website and the state broadband map. Both now provide crowd-sourced data that were used for verification of the Round 3 provider supplied data. Details of those and other verification procedures are provided in the *Verification* section below.

Our speed test activities and the state broadband map both initially created concerns for some members of our provider community. We took great pains to address all those concerns first through informational webinars and then via in-person meetings with individual providers and the NYS Telco and Cable association leadership. We are happy to report that not only were the concerns satisfactorily addressed, but our relationships with the providers and associations were strengthened in the process.

The remainder of this paper describes our methodology for populating the data transfer model and performing data verification; and provides a summary of our data collection results and goals for Round 4.

Provider Participation Summary Tables:

82	**Total Participating Providers**
69	Wireline Providers
14	Wireless Providers (2 are both Wireless & Wireline)
1	Provider is middle-mile only
39	Providers submitted Middle Mile Data

Technology Type	Wireline Census Block Provider Count	Wireline Service Availability by Census Block	Wireline Street Segment Provider Count	Wireline Service Availability by Street Segment	Wireless Provider Count	Wireless Services by Shapefile	Middle Mile Provider Count	Middle Mile Points
Asymmetric xDSL	44	280,976	36	34,205	0	0	25	1,439
Symmetric xDSL	6	62,947	3	176	0	0	0	0
Other Copper Wireline	7	85,730	5	282	0	0	3	131
Cable Modem - DOCSIS 3.0	7	168,469	5	6,778	0	0	0	0
Cable Modem - Other	13	158,933	12	39,451	0	0	2	8
Optical Carrier/Fiber to the End User	21	109,969	15	2,846	0	0	6	576
Terrestrial Fixed Wireless - Unlicensed	0	0	0	0	7	9	0	0
Terrestrial Fixed Wireless - Licensed	0	0	0	0	1	1	0	0
Terrestrial Mobile Wireless	0	0	0	0	6	20	2	13
Other (middle-mile only)	0	0	0	0	0	0	1	2
TOTAL SUBMISSION	**98**	**867,024**	**76**	**83,738**	**14**	**30**	**39**	**2,169**

Populating the Data Transfer Model:

For Round 3, we continued to receive new data from providers in various formats:

1. **Hard copy/.pdf maps**: digitized/georeferenced maps, aggregated availability to the census block and street segment level
2. **Address locations of availability**: geocoded addresses and aggregated to the census block and street segment level
3. **Census block keys and street segment IDs** (Excel worksheets, text files, and shapefiles): where necessary, converted all census blocks to Census 2000 geography and converted TIGERLine streets to New York State street segment geometry
4. **Shapefiles** of wireless coverage areas: added appropriate attribute information where necessary
5. **Kml files –** some fixed wireless and cellular providers supplied coverage area via kml; added appropriate attribute information where necessary

All (non-ESRI) data were converted to ESRI shapefile format with availability aggregated to the census block and street segment level and with provider attributes added (i.e. Provider Name, DBA Name, Technology, Speeds, etc.). Data from the individual shapefiles were loaded into the data transfer model.

Some providers actively report no changes from the previous round. Other (small area) providers were sent maps of their reported Round 2 availability and asked to confirm that there were no changes. In both of these "no change" situations, previously delivered Round 2 data was loaded in the data transfer model. All data were then verified.

The following fields were added to capture NYS specific information:

1. BBConnectionPoint_MiddleMile
   a. Provider Code – internal use
2. BB_Service_CensusBlock
   a. Suffix_Cont – additional field used in street address
   b. Verification – tracks verification results connected with speed test data
   c. Provider Code –internal use
3. BB_Service_RoadSegment
   a. NYS ID – ID code for each street segment
   b. Suffix_Cont – extra field used in street address
   c. Verification – tracks verification results in connection with speed test data
   d. Provider Code –internal use
4. BB_Service_Wireless
   a. Provider Code –internal use

The following domains were extended to allow for plausible and verified attribute values supplied by providers:

1. Other Copper Wireline Down (30) – code 9
   Other Copper Wireline Up (30) – code 9
2. Cable Modem – DOCSIS 3.0 Down (40) – code 10
   Cable Modem – DOCSIS 3.0 Up (40) – code 10
3. Cable Modem – Other Down (41) – code 10
   Cable Modem – Other Up (41) – code 10
4. Terrestrial Mobile Wireless Down (80) – code 7
   Terrestrial Mobile Wireless Up (80) – code 7
5. SPECTRUM USED – code 10 - Other

Verification:

Automated verification was accomplished via the following methods:

1. Business rules built into the 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. Clipping all data to the NYS boundary file

2. Checking for Duplicate: The NTIA Python Script checks for multiple speeds reported by provider & technology on each census block and street segment ID, but it does not check for total duplicate records. In order to check for total duplicate records, we performed the following verification steps:
   a. Created a text column in the street segment and census block feature classes and used the field calculator to concatenate FRN, TRANSTECH, and FULLFIPSID /Street Segment ID on each record.
   b. Summarized the concatenated field to find any records where the COUNT was greater than 1 (indicating a total duplicate record).
   c. Related those >1 COUNT records back to the geodatabase feature class and deleted duplicates.

3. Additional geometry checks: The *Select Layer by Location* tool in ArcGIS was used to check the vertical alignment and area designation of all census blocks loaded in the geodatabase. The parameters of the tool were set to select census blocks in the geodatabase that 'are identical to' a base layer of Census Blocks (2000 vintage) consisting of only ≤ 2 square mile blocks. All census blocks in the geodatabase that vertically aligned with the base layer were selected. The same process was performed on the street segment feature class using a base layer of streets in census blocks greater than 2 square miles.

4. Provider verification: For providers with significant changes from the previous round, we created review maps showing Round 3 availability aggregated to census blocks and street segments. These 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.

5. 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, 4741 records were successfully geocoded and used for verification. Three 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 but technology unknown
    Code 3 = Provider matched but technology is mismatched

    Each census block and street segment availability record involved with this verification activity was assigned one of the above codes. We consider the assignment of these verification codes as the start of our work on the leading practice of establishing record level confidence scale. Further work is planned to create a flexible and informative scale that can be expanded as new data sources and activities are added to our verification workflows.

    b. **FCC speed test records** were used to verify provider reported availability. FCC speed test records lack provider and technology 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 but technology unknown). Here is a statistical summary:

	Number	Percentage
Total Number of FCC Wireline Speed Test Points	59098	N/A
Total Number / Percentage Successfully Geo-coded	30685 / 59098	52%
Total Number / Percentage Successfully IP Searched	21766 / 30685	71%

    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 the availability from

the provider submitted data was confirmed, the next step was to use the provider's own website to determine 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 for that provider from our map. In all cases, the results were reported back to the originator.

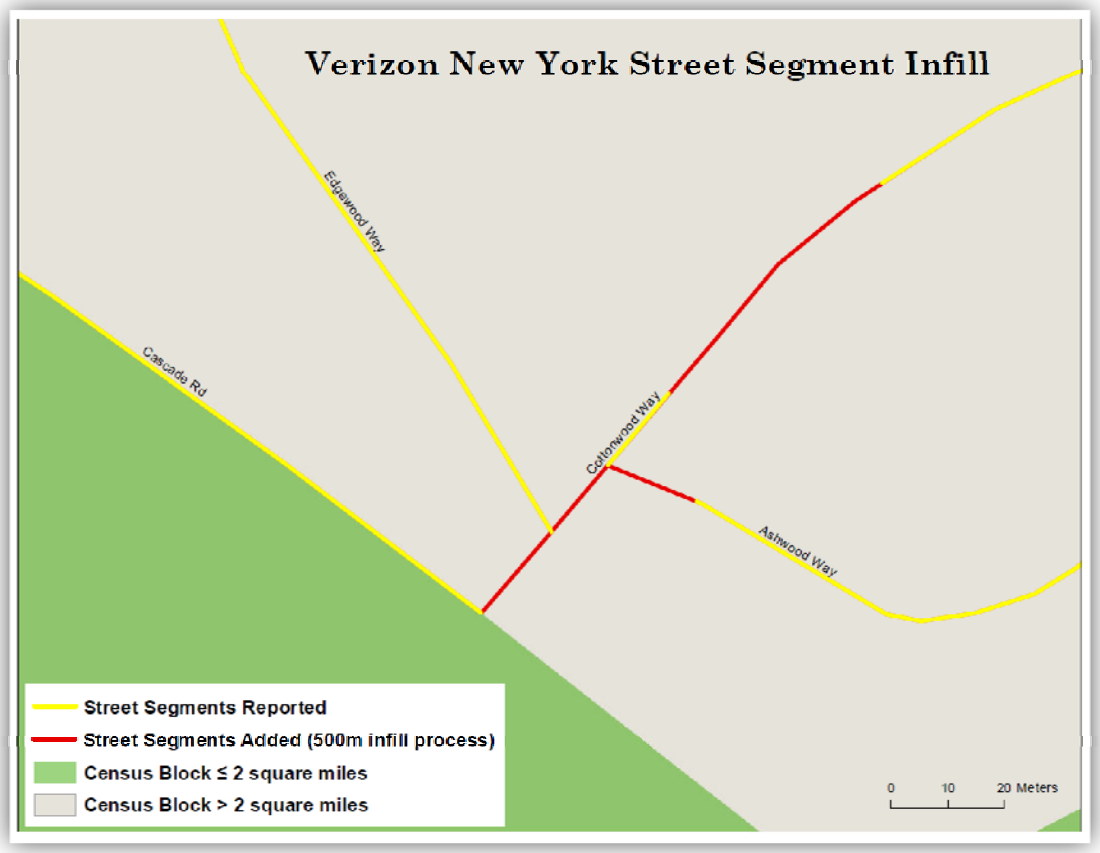
Here are summary statistics for this feedback activity:

- 45 email responses from NYS map
- 37 Blocks, 14 Street Segments questioned
- 9 reported locations were *verified* as having coverage through our process
- No data interpretation/integration errors uncovered: all information that was questioned by public was originally reported by providers
- 40 Blocks or Segments had coverage *removed* after our investigation
- *1 report of no Fairpoint Communication coverage actually resulted in that person getting broadband at their home after Fairpoint's CEO requested they contact him*

6. **Verizon NY (wire-line) specific scrubbing:** Verizon New York submitted data in TIGER/Line street segments 2009 and Census Blocks 2009 as text delimited files.

   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 500 meters or less that fell in between the reported street segments (see image below). Addresses from a sampling of the new street segments were checked through Verizon's website and broadband availability was verified. These segments were added to Verizon's availability and assigned the max advertised speed attributes of the nearest street segment. Street segments that fell outside Verizon's exchange boundary, or fell on a block discontinuous with any other reported blocks were sampled and checked

for availability through Verizon's website and subsequently deleted.



Census blocks were converted to 2000 vintage and reapportioned based on the Census 2000 block area distinction. This posed a challenge since there is not a one-to-one area relationship between 2009 and 2000 blocks. To overcome this discrepancy and maintain accuracy, census blocks that were ≤ 2 square miles in the Census 2009 vintage but over 2 square miles in the Census 2000 vintage were assigned street segments in the portion of the block that corresponded to the smaller 2009 block. Census blocks that were greater than two square miles in the Census 2009 vintage but ≤ 2 square miles in the 2000 vintage were assigned the smaller 2000 block. Another issue occurred where 314 Census blocks reported by Verizon fell outside their exchange boundary. 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 314 census blocks were deleted.

Round 4 Focus:

While increasing the level of participation and improving the completeness and quality of the overall data will continue to be one of our overarching goals, there are some very specific improvement areas that the NYS Broadband Mapping Team will focus on during Round 4. They are:

1. 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

2. Identifying and working with more 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 unserved areas. A number of these small companies that we have contacted thus far have explained that they have very limited human and technical resources and, in many cases, are not able to generate any map-able data on their own. Some have had a consultant generate a propagation model (one time) and that model is now outdated. We intend to work with these providers in order to come up with a solution where we can assist them in mapping their serviceable areas and provide updates as they expand.

3. 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.

4. 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.

5. 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 (discussed above in *Verification* section).

6. 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.