

“White Paper” from *New York* describing Round 2 (Fall, 2010) Data Submission to the NTIA under the SBDD

November 12, 2010

Data Summary

Provider Participation

77 Total Providers

66 Wireline Providers

12 Wireless Providers (2 Wireless providers are both Wireless & Wireline: TVC Albany, Fisher’s Island)

1 Provider is middle-mile only (AT&T Corp); **38** Providers in total submitted Middle Mile Data

Speed Data:

64 Providers submitted speed data at the census block level

2 Providers submitted speed data at the county level

New Data for Round 2:

1. **Received new data** from providers in various formats and converted all 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.):
 - a. **Hard copy/ pdf maps:** digitized/georeferenced maps, aggregated availability to the census block and street segment level
 - b. **Address locations of availability:** geocoded addresses and aggregated to the census block and street segment level
 - c. **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
 - d. **Shapefiles** of wireless coverage areas: added appropriate attribute information
2. **Verified new data with providers:** Created provider review maps showing Round 2 availability aggregated to census blocks and street segments. Providers were given at least five days to respond and initiate any changes or corrections.
3. **Made changes based on provider feedback:** changed aggregated availability or attribute information based on communications with the providers who responded to Round 2 data verification maps. Changes were documented for future reference.

4. Imported final Census Blocks 2000 and New York State Street Segments into the NTIA geodatabase for Round 2 delivery.

25 Providers submitted new data for Round 2

- 1- AT&T Corp- new X, Y coordinates for middle mile only
- 5- Hard copy maps digitized and availability aggregated to the census block & street segment level
- 5- Geocoded address locations aggregated to the census block & street segment level
- 9- Submitted census blocks & street segments
- 5- Wireless, submitted new shapefiles of coverage area

Re-Use of Round 1 Data for Round 2:

1. **Verified data between Rounds 1 and 2** through Provider Review Maps and contact with the individual providers. These providers indicated that their availability had not changed and to reuse Round 1 data for Round 2.
2. Migrated the verified Round 1 data into Round 2 NTIA geodatabase

52 Providers reused their Round 1 data for Round 2

- 19- Hard copy paper maps previously processed
- 13- Geocoded address locations aggregated to the census block & street segment level
- 13- Submitted census blocks & street segments
- 7- Wireless, reused shapefile sent in Round 1

Census Block Records:

Total Census Block records submitted: **795,774**

739,918 (93%) came from **5 large providers:**

- Verizon New York: 219,077
- Covad Communications: 206,028
- CSC Holdings, Inc: 162,412 (81,206 unique)
- Time Warner Cable: 105,361
- Frontier: 47,040

42,986 (5%) came from **16 mid-size providers** (1,004 to 12,388 census block records each)

12,870 (2%) came from **45 small providers** (1 to 790 census blocks each)

Street Segment Records:

Total Street Segment records submitted: **71,327**

60,580 (85%) came from the **5 large providers**

4,954 (7%) came from the **16 mid-size providers**

5,793 (8%) came from the **45 small providers**

Census Block & Street Segment Records Combined:

Total records submitted: **867,101**

800,498 (92.3%) came from the **5 largest providers**

47,940 (5.5%) came from **16 mid-size providers**

18,663 (2.2%) came from **45 small providers**

Wireless Statistics:

Provider Name	DBA Name	FRN	TRANS TECH	Coverage Area (sq mi)	NYS Area (sq mi)*	Percent of NYS Area
Cellco Partnership	Verizon Wireless	0003290673	80	39392.52	47213.79	83.43%
AT&T Mobility LLC	AT&T Mobility	0004979233	80	18717.09	47213.79	39.64%
Sprint Nextel Corporation	Sprint	0003774593	80	16937.37	47213.79	35.87%
T-Mobile USA, Inc.	T-Mobile USA	0006945950	80	3859.52	47213.79	8.17%
Leap Wireless International, Inc. (Cricket Communications)	Leap Wireless International	0002963528	80	2433.88	47213.79	5.16%
Hudson Valley Wireless	Hudson Valley Wireless	0004500591	70	814.46	47213.79	1.73%
Clearwire Corporation	Clearwire	0017775628	80	804.77	47213.79	1.70%
Tivcorp, Inc.	Webjogger Internet Services	0016096083	70	104.12	47213.79	0.22%
Towerstream, Inc.	Towerstream	0007097355	70	7.47	47213.79	0.02%
Fishers Island Telephone Corporation	Fishers Island Telephone	0004324570	70	4.04	47213.79	0.01%
TVC Albany, Inc	TechValley Communications	0006097711	70	1.52	47213.79	0.0032%
Nextlink Wireless, Inc.	Nextlink Wireless	0014286934	71	0.004	47213.79	0.000085%

* U.S. Census Bureau (<http://quickfacts.census.gov/qfd/states/36000.html>)

Mobile Wireless Coverage (Technology code 80):

- **6 Providers**
- **40,224.05** square miles (**85.23%** of New York State)

Fixed Wireless Coverage (Technology code 70, 71):

- **6 Providers**
- **931.61** square miles (**1.97%** of New York State)

Data Integration

Geocoding

The NYS Office of Cyber Security maintains a statewide Streets and Addresses database as part of its Statewide GIS Coordination Program. The data is maintained with edits supplied by local governments across the state which are then supplied to OCS's streets contractor, NAVTEQ. This database provides an exceptionally accurate and up-to-date reference database to support geocoding. OCS geocoded address-based broadband availability data from five providers in Round 2, in addition to 13 providers who supplied address-based data in Round 1 that was verified and resupplied for Round 2.

In the process of geocoding address-based availability data, OCS first uses an address-scrubbing software tool which standardizes the address format and verifies the address from US Postal Service data. We then geocoded these scrubbed addresses geocoding locator files created from our NYS streets and addresses data. We require a minimum of 80% pass rate on the data, but more typically achieve 85% of better match rates. All unmatched addresses are sent back to the provider as part of a feedback loop for correction and resubmission.

OCS notes that we anticipate improved geocoding match rates once ESRI resolves some technical issues which have been logged as trouble reports.

Hardcopy "Sharpie Method"

Many of the small, independent providers have been submitting their broadband availability information as markups of hardcopy maps. OCS created this process for Round 1, based on the suggestion of the New York State Telecommunications Association, which represents many of the small providers. For each provider using this method, we create a pdf map showing the provider footprint and any previously mapped broadband availability information. They typically print this map and mark it up to show streets where they offer service and then send it back to us. Some providers actually mark up the pdf file digitally and send back a revised pdf, but the majority send marked-up printouts.

OCS uses the map to code up census blocks and street segments, with the end product being a new map which is then sent to the provider to verify that we interpreted their markup correctly. In many cases, telephone conversations and email communications take place during the process of transcribing their data into our GIS system. Our GIS staff each handle communications directly with providers as a normal part of their job duties.

Uninhabited Areas

Our data integration processes filter out any census blocks indicated as having wireline broadband availability if those blocks that have been characterized as Uninhabited Areas. Uninhabited Areas are census blocks with zero population, zero housing units, and have been classified as lands where development cannot occur and will therefore not need wireline household broadband. The classifications of uninhabited lands are listed in the table below. The entire census block with zero population and zero housing units was classified as an 'uninhabited area' if the center of the census block fell within one of these areas. Census blocks from the year 2000 were used to determine uninhabited areas. 12,601 census blocks out of 297,494 total census blocks (4.2%) were deemed

'uninhabited' for the Round 2 data submission. This area totaled 2,813.67 square miles of New York State (5.96% of New York's total land area of 47,213.79 square miles).

Uninhabited Land Types	Description
Water	<i>Water areas are uninhabited for Wireline Broadband Availability, but are included for Wireless Broadband Availability</i>
DEC Land	<i>Lands under the care, custody and control of the NYS Department of Environmental Conservation, including Wildlife Management Areas, Unique Areas, State Forests, and Forest Preserve</i>
APA Land Classifications 7, 8, 9, and 10	<i>7- Wilderness, 8- Canoe Area, 9-Primitive Area, 10- Wild Forest; Adirondack Park Agency</i>
State Parks	<i>New York State Parks and Historic Sites</i>
State Recreation Areas	<i>Wildlife Management, Forest Preserves, Reforestation Areas, Recreation Areas, Fish Hatcheries, Canal Parks, and Marine Parks</i>
County Recreation Areas	<i>County Recreation, Parks and Forests</i>
State Campgrounds	<i>State Campgrounds</i>
Municipal Recreation Areas	<i>Parks, Gardens, Recreation Areas, and Forest Preserve</i>
State Non-recreation Areas	<i>State Correctional Facilities, Developmental Centers, Psychiatric Centers, and Military Sites</i>
Federal Non-recreation Areas	<i>Military Sites, National Cemeteries, VA Medical Centers, and an Animal Import Center</i>
Federal Nuclear Sites	<i>Federal Nuclear Sites</i>
Federal Testing Sites	<i>Federal Testing Sites</i>

Use of ETL software

One of the software tools acquired by OCS for this project is Pervasive Software, one of the commercially available Extract, Transform, and Load (ETL) tools. We also acquired the Melissa data option for this software which we used for address scrubbing, as described in the Geocoding section above.

The ETL tool provides for efficient import and load of DBF data from ESRI shapefiles into our Oracle/SDE database. The ETL software is also helpful for identifying duplicate records and for many other data management tasks which would otherwise require SQL scripting.

Data Validation

Crowd-Sourcing

OCS launched the NYS broadband mapping website (www.broadbandmap.ny.gov) in late September, just prior to delivery of Round 2 data. The site provides mechanisms for the public to report errors in the broadband data as shown on the map. While we have received numerous reports of this nature since the site was launched, none occurred prior to our Round 2 delivery. We anticipate being able to report significant use of this crowd-sourced validation technique in future data deliveries to NTIA.

Similarly, OCS began collecting crowd-sourced broadband speed test data in September through our partner the Center for Technology in Government at the University at Albany. We have begun exploring methods to use this information for several levels of data validation, but none of this occurred with data in the Round 2 delivery. We look forward to using this information for validation purposes beginning with Round 3.

Feedback Loops

The Non-Disclosure Agreements (NDAs) that OCS executed with most providers (a few decided to forgo the use of an NDA) includes a 5 day period for providers to review our representation of their data before we publish any maps from it. We have honored this provision by supplying each participating provider with a pdf map depicting their data, including availability, technology, and speeds, as well as their operating footprint. Often the maps result in phone conversations or email exchanges to clarify what is depicted or to note that the map is correct or if there are data problems. If any corrections are made, a revised map is prepared and sent to the provider for a second review.

Business Rules

OCS created a series of business rules for Round 1 which have subsequently been adopted by the NTIA in the geodatabase model. These rules prevent obvious data errors from being loaded into the geodatabase. OCS has supplemented the business rules in the geodatabase model to include additional load checks, which validate data as we load it into the geodatabase in our last step for delivery to NTIA. In addition, we use 6 special business rules which are implemented as SQL queries in Oracle as part of our back-end database management. Any records which fail business rules are sent back to the appropriate provider as errors which they have the opportunity to correct and resubmit.

Validity of Census Blocks

The Round 2 data delivery is coded to Census 2000 blocks. OCS has removed all blocks which occur exclusively over open water, such as the Hudson River or any lakes. If any provider-supplied data occurs within an open water census block, that record is removed and not delivered to NTIA.

OCS also validates that all census blocks in the delivery to NTIA are valid block IDs occurring within the State of New York.

Further filtering of census blocks is done for uninhabited areas, as discussed above.