

# Washington Broadband Mapping

---

## Data Submission Report

*October 1, 2011*

October 1, 2011

---



Sanborn  
320 Miller Avenue  
Suite 80  
Ann Arbor MI 48103

# Washington Broadband Mapping

## Data Submission Report (October 1, 2011)

### Table of Contents

<b>1 INTRODUCTION.....</b>	<b>3</b>
<b>1 OVERALL PROJECT STATUS .....</b>	<b>4</b>
1.1 DATA COLLECTION.....	4
1.1.1 <i>Broadband Data</i> .....	4
1.1.2 <i>Community Anchor Institutions Data</i> .....	6
1.2 DATA PROCESSING.....	7
1.2.1 <i>Conversion Process of Data from Census 2000 to 2010</i> .....	8
Basic 2000 to 2010 Conversion Process: .....	8
1.2.2 <i>Submission 4: Reference Data Creation</i> .....	9
1.2.3 <i>Submission 4: NTIA Submission Data Model Schema Changes</i> .....	10
1.3 DATA VALIDATION .....	10

# 1 Introduction

This report is submitted along with the fourth data submission for the Washington Broadband Mapping Project. This submission includes all data collected so far per the requirements of the National Telecommunications and Information Administration (NTIA) State Broadband Data and Development Grant Program (Docket No. 0660-ZA29) Notice of Funds Availability (NOFA) and formal and informal clarifications to it. Specifically, it includes broadband data collected from broadband providers and community anchor institutions data compiled from various sources for the State of WA. The State of WA has retained a mapping contractor, The Sanborn Map Company to perform the work related to the Mapping Grant for this project. Data from the previous submission is now publicly accessible via the WA Broadband Program (<http://wabroadbandmapping.org/>).

**This document is a supplement to the three previous reports submitted with previous data submissions on May 1, 2010, October 1, 2010, and April 1, 2011 respectively.** Therefore, it builds on the document provided with those submissions. Rather than repeat the contents of the previous report, this document makes incremental updates on various topics where changes have been made in the methodology or reiterates the methodology used. Please refer to the previous documents for further details.

# 1 Overall Project Status

## 1.1 DATA COLLECTION

This section details data collection related to NTIA deliverables which include broadband data and community anchor institution data.

### 1.1.1 Broadband Data

For this submission, Sanborn started data collection efforts on July 13<sup>th</sup> 2011 by sending out data update requests and technical data specifications after NTIA announced all final changes. These were sent to a large list of companies which were compiled from multiple lists (FCC 477 list (dated June 30<sup>th</sup>, 2010), a list provided by the Washington UTC, NTIA's Wireless Internet Service Providers Association (WISPA)) and from any providers that were identified through other sources such as web research, planning meetings, etc. In our technical document, we highlighted the transformation of data from Census 2000 to Census 2010 and given that change, we requested all providers to submit data in the Census 2010 format. Sanborn also uploaded the final data for each provider in NTIA format from the previous submission to the Sanborn Provider Portal. The providers were encouraged to use the provider portal and update their information on it.

We followed the same contact and follow-up protocols as the previous submissions. The following are some of the important changes or no changes:

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 this submission.
2. We continued to not collect data from resellers.
3. We are submitting data for satellites in this submission based on NTIA clarifications. All satellite providers who have provided speed, FRN number and other technology information have been mapped to serve the full state. At present WA data includes only two satellite providers – Hughes and Wildblue. The other satellite providers which operate in WA to the best of our knowledge are Starband and Stratos Offshore Services Co. both of which did not provide adequate attribute information in order to be included on the mapping data.
  - 1) Additionally, given the topography in WA, we have done some Viewshed Analysis to identify areas with no line of sight to satellites. This analysis has been done separately for each provider and for now, the analysis has been done for Wildblue and Hughes satellites. We have provided the resulting data to Wildblue and are working with them to

validate the data at present. Based on what we find, we will decide how to represent such areas of unlikely line of sight on the State Broadband Map.

4. We worked hard to get even more Public Utility Districts (PUDs) in Washington to participate in the program. As previously noted, PUDs are public entities at the County level that lay broadband infrastructure connecting to the end users (i.e. such as fiber to the homes) but WA regulations do not allow them to sell directly to the customers. In these PUD deployments, broadband service is provided by resellers using the infrastructure owned by the PUDs at speeds and costs that the market is capable of bearing. While most PUDs are able to provide their area of availability, it is harder for them to provide the speed bundles that their resellers are using. We held a meeting with NTIA (Andrew McRae) and the State of Washington Program Manager (Will Saunders) to discuss some options for putting PUDs on the map. We held additional meetings with the representatives of several PUDs and made the decision that the map and data should represent the highest speeds that the PUDs are capable of providing should there be a customer willing and able to pay for such a speed. Therefore, the PUD data represents high speeds in rural areas since most of the PUDs are in rural areas and have put in fiber connections to homes. It does not take into account the backbone capacity.
5. Due to NDA restrictions and our inability to accurately flag service by “category of end user”, address points are not included in this submission to NTIA for any commercial provider.
6. Some providers did not submit middle mile elevation or backhaul capacity, particularly when they asked us to reuse previous submission data. Wherever possible, we went back to providers to obtain that information, but it is not available for every record.
7. Terrestrial Mobile Wireless and Terrestrial Fixed Wireless (licensed and unlicensed) were treated as wireless coverage and were delivered as a shapefile. In cases where a provider served using the same technology and spectrum but with different speeds, overlapping areas were removed and the higher speed was assigned.
8. If a cable based wireline provider provides 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.
9. 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.

10. The submission 4 provider data model is currently based on the NTIA data model as of 6/30/11.

We added 5 new providers in this submission – Pogo Zone (terrestrial fixed wireless unlicensed), PUD Skagit (fiber), Northland Cable (cable), Rock Island (wireless) and Tanager (wireline and wireless). In this submission, 59% of the providers submitted new or updated data whereas for 41% of the providers we reused data from their previous submissions. One of the larger wireline providers in WA, Qwest was bought out by Century Link. Also another provider, Megapath merged with Covad. The Qwest and Megapath data are now represented under Century Link and Covad data even though the datasets provided were different. In places where Qwest and Century Link's service overlapped, we used spatial tools to select the highest speeds for a given technology of transmission.

### 1.1.2 Community Anchor Institutions Data

The community anchor institutions data continues to be crowd-sourced through the online data gathering application created by the Sanborn Team. The State of Washington is doing the PR around this data collection and contacting the relevant agencies to request them to fill in data. This has been a slow process and we are getting to a point of diminishing returns with this effort. The numbers of community anchor institutions that have responded so far is provided below:

Category	Name	Total	Total with Broadband Information in Submission 4
1	School - K through 12	2299	1773
2	Library	356	356
3	Medical/healthcare	135	54
4	Public Safety	1706	105
5	University, college, other post-secondary	220	180
6	Other community support - government	343	32
7	Other community support - nongovernmental	344	11

## 1.2 DATA PROCESSING

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 row).
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
  - 1) When a provider submits a service boundary, we select all the blocks and roads inside that shape.
  - 2) If a provider submits a customer address list, the points are geocoded, and then the appropriate block or road segment is selected.
  - 3) If a provider submits block and road information using Census data, we just make sure everything is formatted to the appropriate specifications.
  - 4) If the 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.
  - 5) After each round of processing, we make sure that we only keep unique records. A unique record is defined as having a unique 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 is 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.

6. **Provider Review:** Processed data is all posted to a customized web-mapping tool we commonly refer to as the Provider Portal. All providers were notified once their data was available on the site, and they were always given 3-5 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.**

### 1.2.1 Conversion Process of Data from Census 2000 to 2010

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
  - 1) 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.
  - 1) If less than or equal to 2 --> bring in the 2010 geometry and add that record to the blocks table
  - 2) 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
  - 1) If there is a match, add that record to the roads table
  - 2) If there is not a match, select centroid of existing 2000/2009 segment, and select closest 2010 road
  - 3) 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.
  - 1) 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.)
  - 2) There are some holes in the census crosswalk table
  - 3) 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, then we will add the block to the provider service area.

### **1.2.2 Submission 4: Reference Data Creation**

This section describes the reference data used in submission 4.

#### **BLOCK REFERENCE**

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

1. Block size (AREA) is calculated combining the 2000 land area (ALAND) and water area (AWATER)
2. AREA is converted from square meters to square miles to calculate square mileage (SMI).
3. 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**

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

1. The GT2SMI (Greater Than 2 Square Mile) indicator is set to True when:

- 1) The 2010 road segment is completely within a block that is NOT less than 2 square miles
2. Only minimum and maximum address ranges and a single zip code for each road segment is maintained.

### **1.2.3 Submission 4: NTIA Submission Data Model Schema Changes**

The data model released on June 30, 2011 contained the following changes to the s4 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.

## **1.3 DATA VALIDATION**

Sanborn has continued to perform the same validation on the data as the previous three submissions (details in previous reports). Some minor updates to the validation process are discussed below.

- 1) QC of the data at various steps
- 2) Spatial checks against public and commercial datasets
  - a. For WA, we continued to use the following datasets for validation:
    - i. Exchange Boundaries: for DSL boundaries
    - ii. MediaPrints: for Cable boundaries
    - iii. Speedtest.net data
- 3) Verification by providers
- 4) In this Submission, along with the standard verification by providers using the Provider Portal, we also identified for providers issues that they needed to focus on regarding the findings of our validation team. This also included validation and feedback we received through our website – this submission we have incorporated and integrated several feedback tools in the Interactive Map and information sourced from users is evaluated with respect to provider data and any noted discrepancies are passed back to the provider for correction. In addition, in this round, we incorporated any feedback provided by NTIA for Submission 3. All of these were done by sending providers a letter that identified issues using screenshots and explaining to them

what the error was and then asking them to go fix those errors using the secure provider portal. If providers disagreed with the feedback, we have documented their response.

- 5) Speedtest data collection and other data collection for verification
  - a. We continue to use speedtest data and community anchor data crowdsourced for validation purposes.
- 6) Planning workshops and local validation