

Washington Broadband Mapping

Data Submission Report *2nd Submission*

October 1, 2010



Sanborn
320 Miller Avenue
Suite 80
Ann Arbor MI 48103

Washington Broadband Mapping

2nd Data Submission Report

Table of Contents

1	INTRODUCTION	3
1	OVERALL PROJECT STATUS	4
1.1	DATA COLLECTION	ERROR! BOOKMARK NOT DEFINED.
1.1.1	<i>Broadband Data</i>	4
1.1.2	<i>Community Anchor Institutions Data</i>	6
1.2	DATA PROCESSING	ERROR! BOOKMARK NOT DEFINED.
1.2.1	<i>General Overview</i>	8
1.2.2	<i>Submission 2: Process Modifications</i>	8
1.2.3	<i>Submission 2: Reference Data modifications</i>	9
1.2.3.1	Creation of processing reference data	9
1.2.3.2	Reference data sources	11
1.2.4	<i>Submission 2: NTIA Submission Data Model Schema Changes</i>	14
1.2.4.1	Schema history and evolution	14
1.2.4.2	NTIA data model changes	15
1.3	DATA VALIDATION	17
2	APPENDICES	21

1 Introduction

This report is submitted along with the second 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 Washington has retained a mapping contractor, primed by The Sanborn Map Company for doing all work related to the Mapping Grant for this project.

This document builds on the document provided with submission of data in May. Rather than repeat the contents of the report with Submission 1, this document makes incremental updates on various topics. For this reason, it may be worthwhile to refer to the document submitted with Submission 1 for more details.

2 Overall Project Status

2.1 DATA COLLECTION

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

2.1.1 Broadband Data

For submission 2, Sanborn started data collection on July 1st by sending out data update requests and technical data specifications. These were sent to a large list of companies compiled from FCC 477 list (dated June, 2009) and from a list provided by the Washington UTC. The technical document highlighted the changes from Submission 1 to Submission 2 and requested incremental data. A web conference was also hosted by the WA Department of Information Services (DIS) to explain the changes and respond to any issues or questions from the broadband provider community. Although we sent the technical specifications to all the providers (more than those on the FCC 477 list), we followed up actively with the providers on the 477 list or those who were already participating. This is because most providers outside of the list were found to be non-providers of broadband. We also contacted all providers of wireless broadband available from the Wireless Internet Service Providers Association (WISPA) list.

During this round of the data update, many providers who had refused to participate in the program earlier expressed an eagerness to participate. The main reason for that was the fact that the WA Broadband Interactive Map was already online and they saw the end result of the data and the cost of non-participation to their own business. The public website was a very effective tool in soliciting participation from new providers as well.

In our solicitation for data updates, we told providers that if we didn't hear from them by a certain date, we would default to using their data from Submission 1. We contacted them after the due date a few times but used Submission 1 data if they did not respond.

As with the first submission, we followed the following protocols:

1. We did not collect data from resellers – one company that provided data in the 1st submission (New Edge Holding Company) realized that as resellers, they did not need to participate and withdrew their data.
2. We have not collected data from satellite providers – most were asked for more detailed data but could not provide anything substantial. Based on the details from the Utah study and on further consultation with the State and NTIA, we may allow satellite providers in the next submission.

- 1) Three satellite providers have been identified in Washington – Hughes, Starband, and Wildblue.
3. Affiliates, subsidiaries etc. have been counted as providers. Please note that data for these entities may or may not be reported as a separate FRN if they share the same FRN as their parent company.
4. Most of the rural telcos and WISPs were not able to provide data in the format required by NTIA. Some rural telcos were able to provide their FCC data, boundary information of their coverage area, and some textual description on speeds, technology of transmission, etc. Sanborn accepted their broadband data in whatever format available, then created data products to meet NTIA requirements as closely as possible. The WISPs also provide data in various formats ranging from buffers around last mile locations, to paper maps with circles drawn. In addition, in this submission, we used an online, secure product called the Broadband Provider Portal – this is an interactive site where providers could draw their boundaries, and mark up changes from previous submissions. This provider portal was also used in validation of processed data by the broadband providers.
5. In our efforts to be as complete as possible, Sanborn contacted more providers than those identified on the FCC list of broadband providers, e.g., public providers such as municipal providers and Public Utility Districts. Public Utility Districts in Washington 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 are not allowed to sell directly to the customers. Broadband service is provided by resellers using the infrastructure owned by the PUDs at speeds that the market is capable of bearing. However, given that the program is not collecting reseller data (in some cases there can be more than 20 resellers on a single PUD infrastructure), such areas would go unreported and consequently shown as unserved on the maps. These are also rural areas and areas where other providers are not operating and hence it is critical for the State to map these providers' service area. For this reason, we collected the data from the PUDs and plan on putting them on the map with a note that they would need to visit the PUD site to find out the list of resellers who can provide retail service to them. We have not collected any maximum advertised speeds for these areas but have the technology of transmission.
6. We have not included any data (middle mile) outside of the boundary of the State of Washington.

In addition to the challenges reported on the first data submission report, some additional challenges related to data collection for submission 2 include:

- 1) Change in version of census blocks: The version of census blocks to use and the back-and-forth related to that caused some grief both to us and to the provider community. Most expressed that they cannot change

scripts mid-course and it would be a huge financial burden on them to redo the data and they were not willing to do that. Most were okay to do the switch in the next round provided it is clarified before data collection starts

- 2) Spectrum: The requirement to provide separate polygon for each spectrum was viewed by wireless providers as a change in specifications and none were willing to comply on this request. They believe that the Spectrum information is very confidential and proprietary and not easy for them to share. They expressed that they would like direct notification from NTIA about such changes and that they agreed to provide certain datasets and in certain formats and that this was a change from what they originally agreed to provide.
- 3) Communication with providers: It would help with data collection if NTIA/FCC held an open forum with the providers for changes that are being proposed for that data collection. This should happen before States start data collection and also providing all change information on an NTIA website to the providers so that they are not questioning the credibility of the request from States.
- 4) Public Utility Districts: We need more guidance on how to provide data for Public Utility Districts in Washington. The issue has been discussed above. We have also had a lack of participation from municipal providers as well as other public providers of broadband.

2.1.2 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. The numbers of community anchor institutions that have responded is provided below:

Category	Name	Total	Total with Broadband Information in Submission 2
1	School - K through 12	2031	1437
2	Library	350	345
3	Medical/healthcare	120	9
4	Public Safety	1622	62
5	University, college, other post-secondary	153	179
6	Other community support - government	361	17
7	Other community support - nongovernmental	345	9

tes Broadband Mapping - Home

Please select the institution from the list. **JONES LIBRARY**
 If you do not see your institution on the list, please select 'Other'.

Location Address of Institution (no P.O.Box): 111 E. MAIN
 City: JONES Zip: 73049

Update Address on Map

Mapped Location: (Please be patient while map loads)



If needed, use this tool to place the address point in the correct location on the map.

Does the Institution subscribe to Broadband Service at this location? Yes No

Who is your Broadband Provider? **Au Speed, LLC**

What type of technology is used for your Institution's Broadband Transmission?

What is the DOWNLOAD speed advertised by your Broadband provider?

What is the UPLOAD speed advertised by your Broadband provider?

Are you currently physically located at the Community Anchor Institution address provided above? Yes No

Updated By

Name: morgan

Organization: appgeo

Title:

Phone:

Email: test@appgeo.com

Internet | Protected Mode

Community Anchor Institution: Crowd sourcing Portal

2.2 DATA PROCESSING

2.2.1 General Overview

In general, the submission 2 processes followed the same basic approach that was used in submission 1. The submission 1 process documentation was included with the first submission. The following sections outline the modifications made to the initial processing in order to meet the submission 2 requirements as defined by NTIA.

In summary they can be divided into the following three categories:

- Process Modifications
- Reference Data Modifications
- NTIA Submission Data Model Schema Changes

2.2.2 Submission 2: Process Modifications

Based on NTIA feedback and information provided in NTIA webinar sessions, the submission 2 data processing workflow was changed to support the new NTIA submission requirements:

1. Submission 2 requires a geodatabase with spatial features for all submitted datasets. Initially, submission 1 required text files (as specified in the NOFA) which were to be delivered along with a wireless shapefile in an interim delivery. For submission 2, NTIA has formalized the file geodatabase that we have used.
2. Submission 2 data processing required the use of 2000 Census data for the identification reference for Census blocks. Since we had used 2000 data for Submission 1, this was not a big issue.
3. Per NTIA requirements the source for the roads reference layer is allowed to be the best available source that the state has available. For consistent representation the state road reference data used was Census Tiger Line IDs (TLIDs).
4. Max advertised speed (up/down) should be stored within the blocks, roads, and wireless area datasets.
5. The weighted average speed alternate format will no longer be accepted.
6. Weighted Average Speed is being submitted on a county basis, and was used to populate the new Overview table. However, because we didn't have any data for Maximum Advertized speeds in the Overview class, we had to remove the weighted average speed as well.
7. Due to our NDA restrictions, last mile points will not be submitted to NTIA. Qwest requested that their address points be submitted to NTIA and those are the only points that have been submitted.
8. Wireless coverage should be provided with unique shapes for each spectrum utilized. However, providers with multiple spectrums were not willing to provide data with exact demarcation of the multiple spectrums. Therefore, in order to meet the data model needs, we replicated the polygons for the two spectrums.

9. Terrestrial Mobile Wireless and Terrestrial Fixed Wireless - Licensed were treated as wireless coverage and were delivered as a shape. These types of wireless were not represented in the block and road datasets. We also represent Terrestrial Fixed Wireless – unlicensed with a wireless coverage polygon.
10. All Provider data and Community Anchor Institutions (CAI) locations should be clipped to the state’s boundary. During submission 1, some CAI locations, address points, middle mile points and wireless coverage areas were located just over the border in neighboring states.
11. The submission 2 Provider data model is currently based on the NTIA data model as of September 8, 2010. All proposed changes have been incorporated into the data submitted with this delivery. Any changes are documented in the Provider Data Model Schema changes section of this document and in Appendix 1.
12. Records dropped during data processing will have an associated reason code, and they will not be submitted to NTIA. Dropped records were maintained in a separate similarly formatted dataset and given to the providers so they had an opportunity to correct any issues. Records without required attributes were not submitted to NTIA.
13. The end-user category has been removed from the submission data model for blocks and roads, and is no longer a provider data requirement.

2.2.3 Submission 2: Reference Data modifications

This section describes the reference data schema that will be used during the Reference Data Setup process described later in this document. Reference Data is geometric data used in the NTIA broadband data processing for reporting all collected provider data. Three submission datasets require reference data: block data, road segment data, and overview data. This section provides a description and analysis of the input reference data and a strategy for transforming it into a reference data schema.

2.2.3.1 Creation of processing reference data

During submission 1, provider service delivery data were joined by primary key (i.e. TLID, BLOCKID) to the reference data required by NTIA for submission. Reference data sets were collected and stored by year and type for each state. This raised a number of issues:

- Multiple reference datasets were maintained
- No clear method to define the reference source dataset use for any given feature submitted to NTIA
- In some cases, multiple road data sources were used to improve the road geometry. This increased the time required for analysis.

- Each state processed independently decided projection and reference data specification

Major improvements in geometry accuracy are contained in the 2009 Census block data. Some of the 2009 Census blocks are subdivisions of 2000 Census blocks. These smaller block sizes reflect changes in population, etc. Each of these subdivided blocks adds a single character alphabetic suffix to the 2000 Block ID.

NTIA guidance requested use of 2000 Census data for submission 2, however, the geometry improvements and the addition of new features of the 2009 data suggested that a hybrid dataset using the 2000 id system and the 2009 geometry was used for submission 2 data processing. For the final delivery to NTIA, all geometry was reverted back to the Census 2000 format.

The following is a summary of other key decisions regarding the reference data processing:

- All reference data was combined to form three feature classes for data processing use (i.e. Block, Road Segments, and Overview)
- Only data needed during broadband data processing will be retained (i.e. extra-unused reference columns will be dropped). For instance, State, County, Tract, and Block fields can be generated from the full BlockID field during the publishing process, so these fields are not tracked through the reference file creation process.
- All reference data column names and data types are based on the NSGIC guidance contained in the geodatabase description working paper dated 4/8/10.
- All reference data column names will be prefixed with “ref_” – to indicate to future data processing steps – the data’s origin as reference data (opposed to provider data).

REFERENCE FILE PROJECTION

- WGS_1984_Web_Mercator will be the projection used for all submission 2 reference data processing per NSGIC guidance contained in the geodatabase description working paper dated 4/8/10.

BLOCK REFERENCE

During the block reference file setup, the 2009 BlockID suffix is dropped and the blocks are dissolved (by Block ID) to produce data with 2000 BlockIDs and 2009 shape geometry. This hybrid allows the most recent geometry to be used with provider data that is based on the 2000 census BlockID.

- The ref_CBYear (Census Block Year) column will be set to “HYBR”

- Block size (AREA) is calculated combining the 2000 land area (ALAND) and water area (AWATER)
- AREA 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

Tiger Line IDs (TLIDs), the key column for Census road data, are maintained between the 2000 and 2009 Census data. However, modifications, such as the splitting of a road segment to include a new road intersection, will produce new segments with new TLIDs. One goal of the road reference creation is to make all possible TLID values available for processing. In this example, a provider might use the 2000 TLID for the full segment, or the 2009 TLID for one of the split segments. The combination of the two files may produce duplicate TLIDs, one with 2000 geometry and one with 2009 geometry. To take advantage of the 2009 geometry improvements, when a duplicate TLID is encountered, the one with the 2000 geometry is removed.

- The ref_CBYear (Census Block Year) column is set to indicate the origin year of the road reference data (2000, 2009)
- The GT2SMI (Greater Than 2 Square Mile) indicator is set to True when:
 - o The 2009 road segment is completely within a hybrid block that is NOT LE2SMI (not less than 2 square miles)
 - o The 2000 road segment's centroid is within a 2000 block that is NOT LE2SMI (not less than 2 square miles). The centroid is used because of poor road alignment between the 2000 roads and block.
- Only minimum and maximum address ranges and a single zip code for each road segment is maintained.
 - o In preparation for arithmetic calculations based on address ranges, all extra left and right, as well as address ranges that include alphabetic characters are dropped.

OVERVIEW REFERENCE

Overview data in submission 1 contained three separate feature classes; maximum speed, weighted speed, and pricing data. During submission 1 three separate reference sources (County, CMA, MSA) were also accepted for each of these. In submission 2, all maximum speed data was processed at the block, road, or wireless shape area. Overview was only used to maintain the weighted speed information, and in this submission, only County is accepted as a geography type.

2.2.3.2 Reference data sources

The following data sources were used as reference data sources for submission 2:

BLOCK REFERENCE DATA: 2009 CENSUS BLOCKS

The 2009 Census Block data is the most recent geometry provided by the US Census Bureau and has these characteristics:

- The full Block ID is allotted 17 characters (even though the sum of the component parts only adds up to 16 characters):
 - (2) State
 - (3) County
 - (6) Tract
 - (4) Block
 - (1) Suffix - The 2009 Census Block data allocates a one-character alphabetic suffix to the end of a 2000 Block ID for all blocks that have been subdivided
- Fields of interest include:
 - [BLKIDFP] :: char(17) – Full Block ID
 - [ALAND] :: double(14) – Land Area
 - [AWATER] :: double(14) – Water Area
- The 2009 Census block geometry has been adjusted to correspond with the revised and amended 2009 Census road data.
- This 2009 data represents the new standard for sharing statistical data and is good for matching to the results from our processing. However, historical data may not match this Block ID system.
- This data was downloaded for each state from the following website:

<http://www2.census.gov/cgi-bin/shapefiles2009/national-files>

ROAD REFERENCE DATA: 2000 CENSUS TIGER LINES

The 2000 Census Tiger Line data contains geometry used during the 2000 Census Bureau. The following is a list of characteristics:

- The Tiger Line Identification (TLID) system is stored as a double data type, although it contains only integer values
- Fields of interest include:
 - [TLID] :: double(10) –Originally long integer in TGR file spec (Tiger Line ID)
 - [FEDIRP] :: char(2) – (Feature Prefix Direction)
 - [FENAME] :: char(30) – (Feature Name)
 - [FETYPE] :: char(4) – (Feature Type)
 - [FEDIRS] :: char(2) – (Feature Suffix Direction)
 - [FRADDL] :: double(11) – Originally text field in TGR file spec (From Address Left)
 - [TOADDL] :: double(11) – Originally text field in TGR file spec (To Address Left)
 - [FRADDR] :: double(11) – Originally text field in TGR file spec (From Address Right)
 - [TOADDR] :: double(11) – Originally text field in TGR file spec (To Address Right)

[ZIPL] :: char(5) – (Zip Left)
[ZIPR] :: char(5) – (Zip Right)

- The Census road data is packaged by county. Roads that exist as the boundary between counties will be duplicated in both county files.
- This data has been the standard format for outputting statistical data for the last decade
- This data was downloaded by county as road segments from the following website:

http://arcdata.esri.com/data/tiger2000/tiger_download.cfm

ROAD REFERENCE DATA: 2009 CENSUS TIGER LINES

The 2009 Census Tiger Line data contains the most recent geometry provided by the Census Bureau. The following is a list of characteristics:

- The Tiger Line Identification (TLID) system is stored as a double data type, although it contains only integer values
- Fields of interest include:
 - [TLID] :: double(10) -- (Tiger Line ID)
 - [FULLNAME] :: char(100) – (Full Name)
 - [LFROMADD] :: char(12) – (Left From Address)
 - [LTOADD] :: char(12) – (Left To Address)
 - [RFROMADD] :: char(12) – (Right From Address)
 - [RTOADD] :: char(12) – (Right To Address)
 - [ZIPL] :: char(5) – (Zip Left)
 - [ZIPR] :: char(5) – (Zip Right)
 - [ROADFLG] :: char(1) – (Road Flag – Is segment a road?)
- The 2009 Census Tiger Line road segment geometry was adjusted to correct 2000 segments misalignment; street name, type and directional information were concatenated into one database column (FULLNAME) and new road segments were added.
- The Census road data is packaged by county. Roads that exist as the boundary between counties will be duplicated in both county files.
- This data represents the new standard for sharing statistical data
- This data was downloaded by county as full tiger line data at the following website:

<http://www2.census.gov/cgi-bin/shapefiles2009/national-files>

Source data was filtered by row were [ROADFLG] = yes to create the reference data set.

Note: Where roads were split, because of road alignment correction or new road additions, new reference (TLID) values were assigned to the new road segments by the Census in the 2009 data set.

OVERVIEW REFERENCE DATA: 2009 CENSUS COUNTIES

The 2009 Census County Boundaries are used for reporting of Weighted Average Speed. The following is a list of characteristics:

- The County identification number is stored as a text and allotted 5 characters

- Fields of interest include:
 [COUNTYIDPF] :: char (15) (County Identification Postfix)
 [NAME] :: char (100) (Name)
- This data was downloaded from the following website:
<http://www2.census.gov/cgi-bin/shapefiles2009/national-files>

2.2.4 Submission 2: NTIA Submission Data Model Schema Changes

This section of the document describes the strategy that was used for the development of the specific data schema used for the NTIA submission 2 provider data. The current data model schema is in Appendix 1.

2.2.4.1 Schema history and evolution

In submission 1, NTIA asked the National States Geographic Information Council (NSGIC) to comment and provide a spatial data model that can provide a common format for data submitted to NTIA. The initial NSGIC data model released had a number of issues that needed to be resolved.

NSGIC released the version 2 of the data model close to the submission 1 delivery date. The new model has improved functionality and conforms more closely to the NTIA submission requirements. The NSGIC version 2 model was used as the basis for our internal processing models and for submission 2.

After submission 1, NTIA took ownership of the submission data model, but did not release any changes until mid August. The NSGIC version 2 was used as the basis for our internal processing models. The submission 2 NTIA data model is similar to the NSGIC version 2 model.

To retain as much of the NSGIC v2 /NTIA spatial data model as possible, the relationship between the provider data and the output specification is kept as simple as possible. Here are a few key NTIA submission data model design considerations:

- Submission feature class names reflected the names in the NSGIC v2 specification
- Column data types are based on the NSGIC v2 specification
- Where possible, field names retained the naming conventions of the NSGIC specification
- All road segment address information used the NSGIC specification of a single min, max, zip for each feature
- The data schema for wireless data follows the NSGIC specification for submitting a single feature per spectrum
- To retain Provider Source Information the ID field is calculated as State Name Abbreviation “_”, Short Name. The ID field exists in

- the NSGIC v2 data model, but not the final NTIA submission 2 delivery model. This column is used during processing and was dropped during final processing, prior to submission to NTIA.
- Any Overview records that were not submitted using State-County codes were not delivered.

2.2.4.2 NTIA data model changes

During the processing of provider data for submission 2, a number of issues were raised about the data model requirements proposed by NTIA. A number of specific errors, such as typographical errors in domain values, or inconsistency surrounding processing of null values, etc., were documented and forwarded to NTIA for response. The issues and resolution are included in Appendix 1: NTIA Submission Dataset Schema Changes at the end of this document.

Based on changes made to the NTIA data model, some data processing procedures were required to populate the current NTIA data model. The following is a list of specific data processing changes that have been implemented:

- The following are the rules for removing records for the final NTIA submission:
 - o Basic Assumptions:
 - Remove any record that has a Maximum Advertized speed that did not meet the definition of broadband
 - MaxAdv is only required in wireless
 - MaxAdv can be null in blocks/roads
 - o Criteria for removing records from Blocks/Roads (wireline)
 - Remove records with invalid MaxAdv speeds
 - o Criteria for removing records from Wireless
 - Remove records with invalid MaxAdv speeds
 - Remove records with null MaxAdv speeds
- In addition, the following processing changes are now performed during post-processing and before the final NTIA submission:
 - o There is a new feature class called State Boundary. These shapes were prepped and added to the reference datasets for each state. For NTIA submission output, these were moved into their own feature class.
 - o The Blocks table has the Block ID separately defined as State, County, Tract, and Block ID. The provider data as processed include the full 15-digit FIPS code, which has been parsed to populate these fields.
 - o ID columns no longer exist. They have been dropped from the final processed data.
 - o The Middle Mile, Overview, and Wireless tables all have a field called StateAbbr (2 character alphabetic code). The final publishing script created and populated the StateAbbr field.

- o In the Roads table, the Ref_ Values are used for Street info and Zip Code. Because the processing produces a null value for Ref_City, the City field is populated with Dlv_City.
- o TransTech was converted from string to small integer.
- o Any record with a TransTech value of X, Y, or Z was dropped.
- o Any other field with a value of X, Y, or Z was set to null.
- o Any Elevation with a -9999, -9998, or -9997 was set to null.
- o Any FRN generated during processing (those starting with 00000000__) were converted to a value of 9999.
- o In the Blocks and Roads tables there are new fields called Reseller. Because only data from actual providers was accepted, this field was set to 'No' for all records.
- o Block geometry was converted from hybrid geometry back to 2000 Census geometry.

2.3 Data Validation

Sanborn has continued to perform the same validation on the data as the first submission with some minor updates discussed below.

We believe that validation is a long-term activity that will occur throughout the course of this project and that more validation activities will continue to occur through feedback from the Broadband Interactive Site for Washington, and the State mapping validation and planning workshops throughout the state. For this submission also we completed our initial preliminary checks and validation on the data and identified areas of concern for further investigation and verification.

- At this point, we believe that we do not have sufficient information to alter provider data and we have been careful not to do so unless there are obvious errors such as incorrect block numbers, or unidentifiable street segments, etc.

1) QC of the data at various steps

Sanborn begins with preliminary QC checks on data from the time they are received by us to when they get processed and put through official QC. This continues through various checkpoints during our process (such as looking for the maximum and minimum values, averages of fields, determining what percentage of a field is populated and whether null values are allowed, visual and spatial checks, etc.). This process has been applied to all datasets received from providers. Because the broadband data provided to us have varied widely in completeness and formats, these checks have been challenging to keep consistent and uniform without additional and repeated checks by Sanborn.

When incomplete data is submitted, Sanborn has made efforts to get correct data by going back to the providers. Some providers have given explanation of missing data, or resubmitted data. Others have responded that they provided what they could, or have not responded at all. Some of the information submitted to NTIA includes incomplete data where certain fields of information were not filled in by the providers.

2) Verification by providers

In this submission, Sanborn used a secure Broadband Provider Portal to upload all processed provider data for review. Markup tools were provided in this application to make any corrections. Through this tool, Sanborn has provided original data, processed data, and some background layers for providers to visualize their service areas and speeds across the service areas. Instruction documents and videos are also available for users to understand requirements. In some cases, new users have used this tool to provide us the submission data. Most providers were given at least 4 business days for reviewing the data and

providing us any feedback. The feedback resulted in some requests for data corrections, which were made by Sanborn. Not all providers responded with feedback on their maps. In such cases we assumed that they were in agreement with their area delineation shown on the maps we provided them.

The business rule changes provided by NTIA that restricted speeds for different technology types resulted in some areas being dropped for providers after they had done validation. We made every attempt to check if they were correct in their speed reporting and most came back with a response that they are okay us dropping their records if NTIA was not accepting higher speeds. Others came back reluctantly and agreed to bring their speeds down to allowable ranges so that their data would not be dropped completely.

Some providers continue to be unhappy with the mapping to blocks and street segments indicating that their broadband coverage went beyond their service areas. Sanborn explained this was due to the compliance requirements per the NOFA , that census blocks had to be wholly reported to be served or not served and could not be clipped back to only their service area. By asking provider to validate draft maps of their service areas Sanborn was able to improve the data for the providers significantly. It also helped to persuade providers to participate in the program because they knew they would have the opportunity to correct any mistakes or miscommunications.

3) Spatial checks against public and commercial datasets

Sanborn continues to use the following publicly or commercially available datasets to check for areas of potential concern:

Exchange Boundaries: for DSL boundaries

MediaPrints: for Cable boundaries

Speedtest.net data

We are looking into purchasing a new version of Media Prints that comes with Fiber data and can be used for validation purposes. This will be used for Submission 3.

4) Speedtest data collection and other data collection for verification

Sanborn has continued to collect speed test data through our Community Anchor Institution survey form online and from the public at large. We are waiting for a critical mass to create speed maps for the state, we can compare the results against information we have from providers.

In addition to this, Sanborn asked Community Anchor Institutions to give information on providers they subscribe to. This data has been collected and is being used for verification purposes.

5) Planning workshops and local validation

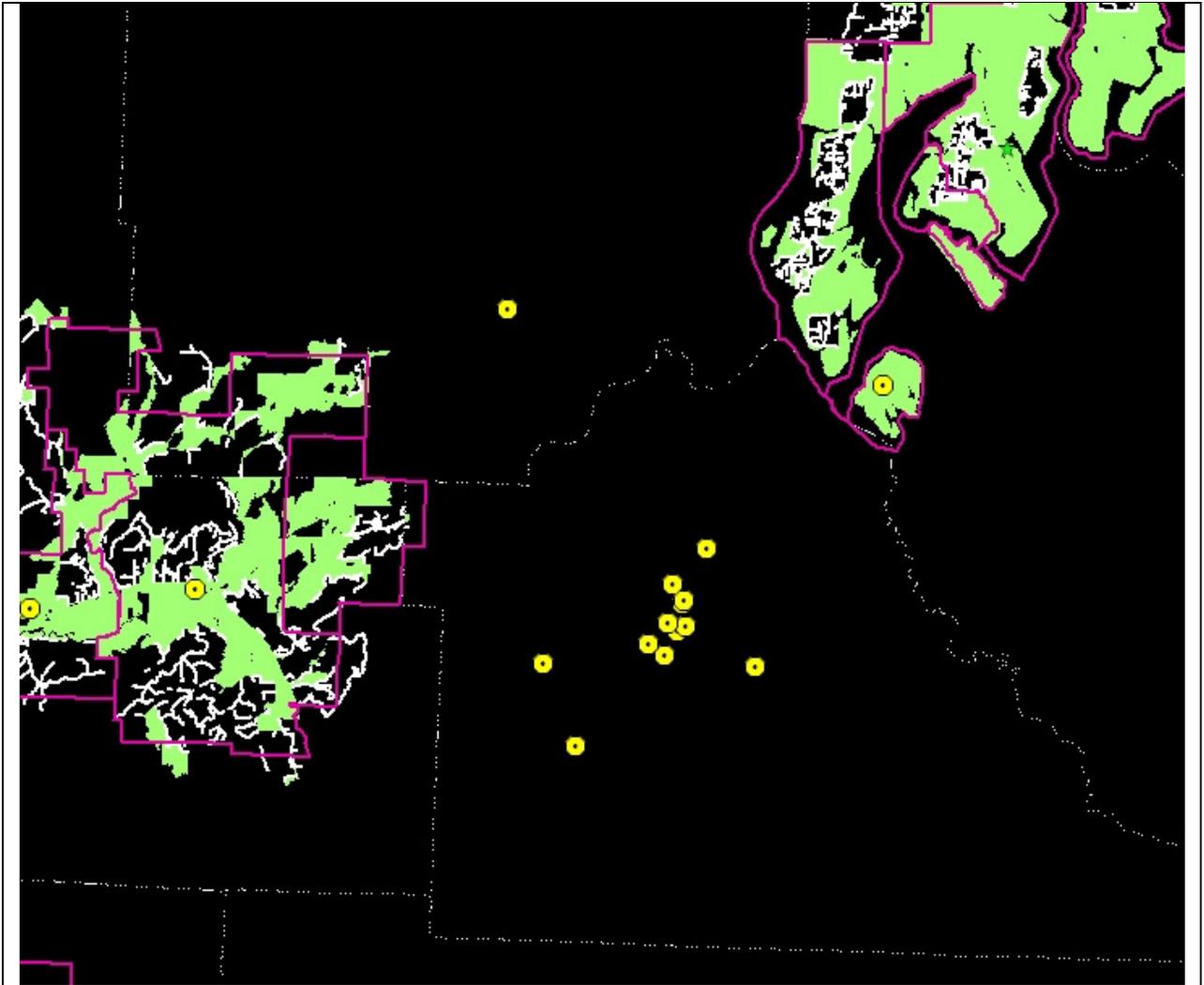
In the month of September, DIS conducted several planning workshops all across the state that were being attended by local officials, broadband providers, representatives from the education and higher education, hospitals, businesses and the community at large. These workshops are soliciting response on the maps. However, given the timing of these workshops, all corrections on the maps will be provided back to the broadband providers for the feedback loop and corrections made in the third deliverable.

2.3.1 Data Validation Conclusions

Data validation continues to be a challenging activity. There is no complete truth sometimes and different pieces of evidence are collected and pieced together to point discrepancies that are explored in more detail. Commercial datasets are often self-reported by the companies and subject to the same errors that we get from providers directly, and sometimes exaggerated by the fact that there are different vintages and resolution and hence the comparison is not easy. Speed test locations are also sometimes incorrect and similar issues exist with all crowd-sourced data.

For example, in validating one of the providers in WA using exchange boundaries, and community anchor institutions, we found evidence of some issues as shown in the screenshot below. The provider's data (in green blocks and white roads) was more or less within the exchange boundary (in magenta) but the community anchor data (shown as yellow and black dots) were telling us that they receive service from this provider in an area that was neither on the exchange boundary, nor on the provider data. Further investigation revealed that this provider provides service to a school district under a special agreement and cannot service those outlying areas in a retail capacity to other users. Therefore, no changes in the data were needed. This example shows how there is no absolute truth exists and that data validation cannot change data arbitrarily based on only one evidence or two.





Discrepancies between provider data, publicly available data and crowd-sourced data.

3 Appendices

Appendix 1: NTIA Submission Dataset Schema Changes

The September 8th version of the NTIA data model was used as the basis for the submission 2 data model. This appendix describes the changes made to this data model due to issues that were found. Based on NTIA analysis, the issues fell into three categories:

1. "All Other" TRANSTECH code - The All Other category was intended for cases where the technology did not fit any other of the existing ones in the model and no business rules have been applied to it. The valid range of speeds used for this should be the ones listed in the Federal Register although some users have already reported to us higher available speeds.
2. Null Values for Overview MAXADDOWN and MAXADUP - Nulls are not allowed on the Overview feature class for MAXADVUP and MAXADOWN by design. The overview feature class represents a higher level or more generalized type of data, and null information for speeds is intentionally not allowed at that level. For more detailed layers like census blocks, road segments, and addresses these fields can have null values. As a result, no Overview records were submitted sine all MaxAdv speeds were reported in blocks and roads. It should be noted that Weighted Average Speed was collected, but not delivered, for this reason.
3. Coded Value Domains - Some of the coded value domains are not linked to the correct subtype, these can be fixed in the database. Many of the speed codes are the same across transmission technologies minimizing the overall impact. The subtype links will be corrected in the next version of the model. The table below lists the data model issues that were found, and the resolutions (i.e. changes to the data model) that were made:

#	Feature Class	Issue	Resolution
1		TRANSTECH of All Other does not have any sub domain tables to select	Created two domain tables, one called All Upload (2 -11) and another called All Download (3-11)
2		Domain Transmission of Technology has TRANSTECH 40 is named Cable Modem - DOCIS 3.0 Down	removed the word "Down" - new values is DOCIS 3.0
3	BB_Service_Address	Subtype for TransTech Symmetric xDSL - MAXADDOWN domain set to AxDL Down	domain set to Symmetric xDSL Down
4	BB_Service_Address	Subtype for TransTech Symmetric xDSL - MAXADUP domain set to AxDL Up	domain set to Symmetric xDSL Up
5	BB_Service_Address	Subtype for TransTech Other Copper Wireline - MAXADDOWN domain set to Satellite Down	domain set to Other Copper Wireline Down
6	BB_Service_Address	Subtype for TransTech Other Copper Wireline - MAXADUP domain set to Symmetric xDSL Up	domain set to Other Copper Wireline Up
7	BB_Service_Address	Subtype for TransTech Other Copper Wireline - TYPICDOWN domain set to Symmetric xDSL Down	domain set to Other Copper Wireline Down
8	BB_Service_Address	Subtype for TransTech Other Copper Wireline - TYPICUP domain set to Symmetric xDSL Up	domain set to Other Copper Wireline Up

#	Feature Class	Issue	Resolution
9	BB_Service_Address	Subtype for TransTech TRANSTECH 40 is named Cable Modem - DOCIS 3.0 Down	removed the word "Down"
10	BB_Service_Address	Subtype for TransTech Terrestrial Fixed Wireless - Unlicensed - TYPICDOWN domain set to Terrestrial Fixed Wireless Licensed Down	domain set to Terrestrial Fixed Wireless Unlicensed Down
11	BB_Service_Address	Subtype for TransTech Terrestrial Fixed Wireless - Licensed - MAXADOWN domain set to Terrestrial Mobile Wireless Down	domain set to Terrestrial Fixed Wireless Licensed Down
12	BB_Service_Address	Subtype for TransTech Terrestrial Fixed Wireless - Licensed - MAXADUP domain set to Terrestrial Mobile Wireless UP	domain set to Terrestrial Fixed Wireless Licensed Up
13	BB_Service_Address	Subtype for TransTech Terrestrial Fixed Wireless - Licensed - TYPICDOWN domain set to Terrestrial Mobile Wireless Down	domain set to Terrestrial Fixed Wireless Licensed Down
14	BB_Service_Address	Subtype for TransTech Terrestrial Fixed Wireless - Licensed - TYPICUP domain set to Terrestrial Mobile Wireless Up	domain set to Terrestrial Fixed Wireless Licensed Up
15	BB_Service_Address	All Other - MAXADDOWN - no domain set	domain set to All Download
16	BB_Service_Address	All Other - MAXADUP - no domain set	domain set to All Upload
17	BB_Service_Address	All Other - TYPICDOWN - no domain set	domain set to All Download
18	BB_Service_Address	All Other - TYPICUP - no domain set	domain set to All Upload
19	BB_Service_CAInstatutions	TRANSTECH 40 is named Cable Modem - DOCIS 3.0 Down	removed the word "Down"
20	BB_Service_CAInstatutions	All Other - MAXADDOWN - no domain set	domain set to All Download
21	BB_Service_CAInstatutions	All Other - MAXADUP - no domain set	domain set to All Upload
22	BB_Service_CensusBlock	TRANSTECH 40 is named Cable Modem - DOCIS 3.0 Down	removed the word "Down"
23	BB_Service_CensusBlock	Cable Modem - DOCIS 3.0 - MAXADDOWN does not have any domain assigned	domain set to Cable Modem - DOCIS 3.0 Down
24	BB_Service_CensusBlock	Cable Modem - Other - MAXADDOWN domain set to Cable Modem DOCIS 3.0 Down	domain set to Cable Modem - Other Down
25	BB_Service_CensusBlock	Cable Modem - Other - MAXADUP domain set to Cable Modem DOCIS 3.0 Up	domain set to Cable Modem - Other Up
26	BB_Service_CensusBlock	Cable Modem - Other - TYPICDOWN domain set to Cable Modem DOCIS 3.0 Down	domain set to Cable Modem - Other Down
27	BB_Service_CensusBlock	Cable Modem - Other - TYPICDUP domain set to Cable Modem DOCIS 3.0 Up	domain set to Cable Modem - Other Up
28	BB_Service_CensusBlock	TRANSTECH 71 is named Terrestrial Mobile Wireless	renamed to Terrestrial Fixed Wireless - Licensed
29	BB_Service_CensusBlock	All Other - MAXADDOWN - no domain set	domain set to All Download
30	BB_Service_CensusBlock	All Other - MAXADUP - no domain set	domain set to All Upload
31	BB_Service_CensusBlock	All Other - TYPICDOWN - no domain set	domain set to All Download
32	BB_Service_CensusBlock	All Other - TYPICUP - no domain set	domain set to All Upload
33	BB_Service_RoadSegment	Asymmetric xDSL - MAXADDOWN domain set to AxDSL Up	domain set to AxDSL Down

#	Feature Class	Issue	Resolution
34	BB_Service_RoadSegment	Asymmetric xDSL - MAXADUP domain set to AxDSL Down	domain set to AxDSL Up
35	BB_Service_RoadSegment	Asymmetric xDSL - TYPICDOWN domain set to AxDSL Up	domain set to AxDSL Down
36	BB_Service_RoadSegment	Asymmetric xDSL - TYPICUP does not have any domain assigned	domain set to AxDSL Up
37	BB_Service_RoadSegment	Terrestrial Fixed Wireless - Unlicensed - TYPICDOWN domain set to Terrestrial Mobilewireless Down	domain set to Terrestrial Fixed Wireless Unlicensed Down
38	BB_Service_RoadSegment	All Other - MAXADDOWN - no domain set	domain set to All Download
39	BB_Service_RoadSegment	All Other - MAXADUP - no domain set	domain set to All Upload
40	BB_Service_RoadSegment	All Other - TYPICDOWN - no domain set	domain set to All Download
41	BB_Service_RoadSegment	All Other - TYPICUP - no domain set	domain set to All Upload
42	BB_Service_Wireless	Cable Modem - DOCIS 3.0 - MAXADDOWN domain set to Other Copper Wireline Down	domain set to Cable Modem - DOCIS 3.0 Down
43	BB_Service_Wireless	Cable Modem - DOCIS 3.0 - MAXADUP domain set to Other Copper Wireline Up	domain set to Cable Modem - DOCIS 3.0 Up
44	BB_Service_Wireless	Cable Modem - DOCIS 3.0 - TYPICDOWN domain set to Other Copper Wireline Down	domain set to Cable Modem - DOCIS 3.0 Down
45	BB_Service_Wireless	Cable Modem - DOCIS 3.0 - TYPICUP domain set to Other Copper Wireline Up	domain set to Cable Modem - DOCIS 3.0 Up
46	BB_Service_Wireless	Cable Modem - Other - MAXADDOWN domain set to Other Copper Wireline Down	domain set to Cable Modem - Other Down
47	BB_Service_Wireless	Cable Modem - Other - MAXADUP domain set to Other Copper Wireline Up	domain set to Cable Modem - Other Up
48	BB_Service_Wireless	Cable Modem - Other - TYPICDOWN domain set to Other Copper Wireline Down	domain set to Cable Modem - Other Down
49	BB_Service_Wireless	Cable Modem - Other - TYPICUP domain set to Other Copper Wireline Up	domain set to Cable Modem - Other Up
50	BB_Service_Wireless	All Other - MAXADDOWN - no domain set	domain set to All Download
51	BB_Service_Wireless	All Other - MAXADUP - no domain set	domain set to All Upload
52	BB_Service_Wireless	All Other - TYPICDOWN - no domain set	domain set to All Download
53	BB_Service_Wireless	All Other - TYPICUP - no domain set	domain set to All Upload
54	BB_Service_Wireless	All Other - STATEABBR - no domain set	domain set to STUSPS
55	BB_Service_Overview	TRANSTECH 40 is named Cable Modem - DOCIS 3.0 Down	removed the word "Down"
56	BB_Service_Overview	Terrestrial Fixed Wireless - Unlicensed - MAXADUP domain set to Terrestrial Fixed Wireless Licensed Up	domain set to Terrestrial Fixed Wireless Unlicensed Up
57	BB_Service_Overview	All Other - MAXADDOWN - no domain set	domain set to All Download
58	BB_Service_Overview	All Other - MAXADUP - no domain set	domain set to All Upload

In addition to the items above, the following changes were made based on NTIA recommendation:

- 1) Allow terrestrial fixed wireless (licensed and unlicensed) upload speeds of 2 (add this as a valid value in each of the terrestrial upload domains)

- 2) In the wireless spectrum domain, “change is Unlicensed (including broadcast television “white spaces”) Spectrum Used to provide service.” To “is Unlicensed (including broadcast television “white spaces”) used to provide service”
- 3) Based on the 9/29 NTIA webinar, all records that meet the definition of broadband are being kept in the data. However, the data model was not changed, so many speed values will show up as digits (i.e. 2, 9, 10 ,11) rather than text domain descriptions.
- 4) For submission 1 WGS 84 Web Mercator projection was used. For submission 2, we are projecting the NTIA submission data to the required WGS_84 projection.