Montana Broadband Mapping Methodology Report

Submitted To:
Chad Hultin
Montana Broadband Mapping Project Coordinator
Montana Department of Commerce

Submitted By:
Fred Gifford
Tetra Tech EC Inc.
and
Ken Wall
GeoData Services Inc.
March 24, 2011

Executive Summary

The following report describes methods and issues related to the April 1, 2011 deliverables to NTIA for Broadband Mapping in Montana. This data submission is compliant with all guidance and specifications provided by NTIA as of March 24th, 2011. As per NTIA guidance we are using the Jan 13th, 2011 version of the Broadband data model and the March 14th, 2011 version of the validation script.

Montana has developed a very robust operational data model, components of which are described in this report, to support our broadband mapping efforts. We feel our operational model can support any reasonable modifications to NTIA requirements or the NSGIC data model. Since this deliverable format is derived from our operational data model, we anticipate some modifications will be required. We are able to take best practices recommendations from the NTIA and incorporate those into the final deliverable without major modifications of our work flow and operating rules.

Our mapping process starts with infrastructure points (central offices, remote terminals, wireless towers and antenna locations, middle mile and backhaul), cable franchise areas, and anchor institution addresses. When providers have not supplied detailed information of their service areas that can be mapped at the census block level, coverage models are derived dynamically from this infrastructure based on geoprocessing techniques specific to each broadband technology. Examples of geoprocessing techniques include developing propagation models using the Longley-Rice model for wireless coverage and using infrastructure points in conjunction with the road network to predict the area served for DSL coverage.

We have developed a system to quantify "validated" data for the purpose of determining what is suitable for delivery to NTIA. The operational data model maintains reliability and validity codes, together with completeness checks to track which data elements are complete or still in process of refinement. Infrastructure is compared to public data, independent measurements, and telecommunications provider submittals at varying levels of geography. As more data is obtained from providers and systematically checked against infrastructure points, the reliability and validity progress from 1 (not validated or reliable) to 10 (validated and reliable). Completeness is primarily dependent on provider input, and can be supplemented in many instances with independent measurements. The process is iterative. Five of the providers included in this data set submitted infrastructure data at the address level. The remainder have submitted at a coarser geographic scale, most often by census tract, small scale paper or digital map, or generalized town location. Our validation methods provides the ability to use general information and iteratively cross check and improve the coverage models as more accurate data is obtained.

Reliability, Validity and Completeness

Reliability codes apply to the source data points and polygons and assess the authority of the source we obtained the data from and the level or coarseness of the geography (address or town). Validity codes are determined from cross checks of data sources and the number of independent sources of verification. These are as simple as comparing speed test locations against DSL modeled polygons, or as complex as geospatial analysis operations such as a kernel density function cluster analysis. Completeness is determined by public sources, independent measurements or provider submittals and checks on the domain classes required for the final NTIA deliverables such as Technology of Transmission domains, Speed Test domains and serving facility and wireless spectrum facility types and categories. The categories for these, and the subsequent records in our operational geodatabase tables grow and change as new data is obtained. We are maintaining these as feature level metadata tied to points and polygons maintained by analysts and technicians in a wiki table and coding them to the geodatabase. In this way the unique situations that arise can be cataloged and maintained with some level of flexibility while contributing to the final indices in a controlled fashion.

Reliability Codes

The two factors incorporated in reliability codes include the level of geography that was used as a source or provided as a clarification of location and the authority of the source for the information. We are also considering clusters of point information from independent measurements and sources to be higher in reliability than individual point information.

Generally, the coarser the source geography the lower the resultant score. Everything besides an address or street intersection, latitude/longitude location, or location provided in a georeferenced digital source is assigned a reliability score less than 5. This applies to source data coming (e.g. a central office located in a city instead of an address) and review comments on a previously mapped location (e.g. “That location is wrong, I know it is on the south side of town”).

We have incorporated the reliability code into our last point of aggregation (LPA) and provider coverage geodatabase files, and into some of the publicly available data (PAD) geodatabases. We are also carrying a short text field (50 characters) with a descriptive rationale for the score. This will allow us to focus more on the lower scores that need to be confirmed, and ignore the high confidence data scored as 9’s and 10’s.

Reliability Codes		
Code	**Description**	**Detailed Description**
0	Not assigned	• Not yet assigned
1	Level 1	• Checked but unverified
2	Level 2	• County • Presence by other coarse geography (e.g. administrative region)

3	Level 3	• City • Census tracts • Cable Plus (area likely to have been annexed into an incorporated town or CDP)
4	Level 4	• Cable - incorporated • Zipcodes • Census blocks
5	Level 5	• GeoTel unverified • Confirmed by provider or anchor institution key advisor but to geography coarser than address or intersection
6	Level 6	• Qwest/Midcontinent or other web site random testing check • Speed test from individual average residential
7	Level 7	• From anchor institution key advisor Webex • GeoTel verified address only with no 3rd party confirmation from public sources o Building unverified • Speed test from anchor institution
8	Level 8	• From provider • FCC ULS or ARS • Geotel verified address and possibly verified by 3rd party source (Google Streetview) o Another provider's sign is on building (usually Qwest) • Geotel possibly verified by 3rd party source (NAIP, Google Streetview) • From state authoritative public data source (e.g. DCN or SummitNet) o Address or building unverified • Speed test from cluster of average residential
9	Level 9	• From provider with authoritative confirmation • Geotel verified address and verified by 3rd party source (NAIP, Google Streetview) o Provider sign on building o Tower or dish visible • From provider or anchor institution check of our data * Root Wireless
10	Level 10	• From 2+ authoritative confirmations

Validity Codes

We include validity codes in the last point of aggregation infrastructure data which drives creation of the DSL models. We also include validity codes in each of the final technology of transmission deliverables for polygons and point feature classes. The scales of validity vary by each major type and function.

Infrastructure Validity Codes

The purpose of this validity code is twofold:

1. To determine which infrastructure points are turned into DSL model coverages
2. To use as a reference in other coverage validity checks

Infrastructure Validity Codes		
Code	**Description**	**Detailed Description**
0	Level 0	• Not yet assigned
1	Level 1	• Not yet assigned
2	Level 2	• Not yet assigned
3	Level 3	• Checked against MT PSC Report or DSLReports at the town level • Checked against SummitNet anchor institution data
4	Level 4	• Checked against two or more independent public sources at the town level • Checked against provider public data (e.g. Qwest ICONN) at the town level
5	Level 5	• Not yet assigned
6	Level 6	• Confirmation of DSL or cable from authoritative public data to broader geography than address not confirmed by provider
7	Level 7	• Authoritative public data at address level (e.g. Geotel) not confirmed by provider
8	Level 8	• Provider submission at the census tract level • Provider website independent address checks (Qwest, Verizon)
9	Level 9	• Provider submission at the census block level
10	Level 10	• Provider submission at the address level

Final Technology of Transmission Validity Codes

The purpose of this validity code is twofold:

1. To determine which elements are loaded in the spreadsheet provider submission packages in their review
2. To determine which provider coverages are chosen for submittal with one of the NTIA deliverables (April 15, June 24)

Final Technology of Transmission Validity Codes		
Code	**Description**	**Detailed Description**
0	Not assigned	• Not yet assigned
1	Level 1	• Unassigned at this time
2	Level 2	• Unassigned at this time
3	Level 3	• Checked against MT PSC Report or DSLReports at the town level • Checked against SummitNet anchor institution data
4	Level 4	• Checked against two or more independent public sources at the town level • Checked against provider public data (e.g. Qwest ICONN) at the town level
5	Level 5	• Confirmation of DSL or cable from authoritative public data
6	Level 6	• Provider website independent address checks (Qwest, Verizon) • Provider submission at the census tract level

7	Level 7	• Provider submission at the census block level • Provider submission at the census block level confirmed by Speed test cluster OR RootWireless
8	Level 8	• Provider submission at the address level
9	Level 9	• Provider submission at the address level confirmed by Speed test cluster OR RootWireless
10	Level 10	• Provider submission at the address level confirmed by Speed test cluster OR RootWireless

State Specific Issues

The most notable issue specific to Montana is the lack of non-disclosure agreements (NDAs) with the providers. To date no provider has agreed to sign a NDA in Montana due to open records laws in the State. However, the vast majority of broadband providers in the State have elected to cooperate with the project and have provided at least some information about their coverage areas. Where providers have not provided data, or not provided adequate data we have used a variety of methods including modeling, field mapping, and use of public sources to develop map data.

Detailed Processing Steps

BROADBAND COVERAGE (BB_SERVICE_CENSUSBLOCK AND BB_SERVICE_ROAD_SEGMENT)

LOGICAL CONSISTENCY REPORT

Data submitted by broadband providers was accepted as is when it was provided as a broadband coverage or at a census block level. Provider coverage submitted at a coarser geographic scale was supplemented with public data, independent measurements and GIS modeling techniques. When independent measurements were available for typical broadband speeds and modeled location of infrastructure, some provider data was overridden or supplemented.

COMPLETENESS REPORT

All data was submitted by broadband providers was mapped in complete form, except where independent measurements were used to supplement provider submittals. Several providers did not participate in the broadband mapping process, including some that were suspected to be providers.

PROCESS DESCRIPTION

Process final broadband coverage from provider submissions.

Broadband providers that chose to submit coverage data did so in a wide variety of formats, levels of completeness, and at varying geographic scales including: narrative descriptions, analog and digital coverage maps, CAD files, GIS shapefiles and geodatabases, KMZ and KML files, FCC 477 reports, and data spreadsheets. All data formats were accommodated and processed whenever possible.

If data was submitted by a provider in a format that did not allow mapping at the census block level of geography, providers were sent standardized maps that included census blocks and a data spreadsheet in an attempt to standardize the inputs and increase the geographic granularity of the provider data submission.

Although each provider had individual characteristics and nuances in their data submissions, several data patterns can be described generalizing the provider submissions.

FCC FORM 477 REPORT OR SIMILAR FORMAT:

Broadband providers are required to submit FCC Form 477 reports twice a year to the FCC; recently 477 submissions have been done using a structured web site maintained by the FCC. The 477 reports require broadband providers to submit a list of census tracts with the number of subscribers based on maximum advertised downstream and upstream speed tiers. Several providers submitted their actual FCC 477 report or a modified version in analog or digital format.

HOW THEY WERE HANDLED:

FCC Form 477 reports were entered into a standardized format that included the census tract ID code, maximum advertised downstream and upstream speed tier code, and number of subscribers if available. Since the FCC 477 reports requires providers to submit data for all speed tiers within a census tract, only the highest maximum advertised speed for any given census tract was entered into the standardized spreadsheet in order to be compliant with the definition of broadband service. The spreadsheets were then joined to a census tract feature class template that included the attribute fields from the NTIA schema. The resulting feature class was a geographical representation of the FCC 477 report including the technology of transmission and speed information. This feature class was used in conjunction with validated LPA data to run the DSL or Cable geoprocessing models respectively. The resulting census block selection from the DSL or Cable models were added to a standardized review map and returned to the provider for confirmation. When providers returned additional data at a finer level of geographic detail they were processed as census block or coverage provider data as described below. For those that did not respond, the final submission was our best modeled estimates of their coverage at the census block level for DSL and/or Cable technologies. Providers that submitted FCC 477 data for fiber to the end user or fixed wireless could not be mapped and were not included in the final broadband map unless they provided additional data.

PROVIDERS SUBMITTING CENSUS BLOCK COVERAGE:

Census blocks submitted by providers representing their broadband coverage area come in a wide range of formats including: analog and digital maps, CAD files, GIS shapefiles and geodatabases, tabular lists, and spreadsheets.

HOW THEY WERE HANDLED:

All census block submittals were loaded into a census block feature class template that included all of the attribute fields from the NTIA schema. Census 2000 geography was used as required by NTIA. Domain codes were entered in the appropriate attribute field for technology of transmission, maximum advertised downstream speed, and maximum advertised upstream speed. If a provider did not identify the technology of transmission for any given census block or blocks, they were contacted by phone or email in order to obtain this information. In instances where speed information was not included in the data submission providers were contacted and asked to supply this data; in cases where the provider refused to give either the downstream, upstream, or both speeds, the lowest domain code was entered in the applicable attribute field. Standardized confirmation maps were created for each provider by type of technology and sent to the provider for review. Once processing was completed for a provider's census block submission, they were run through an ESRI geoprocessing model that performed several quality control-quality assurance tests and selected census blocks less than or equal to two square miles and road segments that intersected with census blocks greater than two square miles.

PROVIDERS SUBMITTING COVERAGE DATA:

Provider submitted coverage data were differentiated from the other types of geographic data submissions coarser than a census block since they represented the full and explicit range of broadband coverage. Similar to the other types of data submissions, coverage data also came in a wide range for formats including: analog and digital maps, CAD files, GIS shapefiles and geodatabases. Coverage data was submitted by several providers or was available on several providers' websites.

HOW THEY WERE HANDLED:

All coverage data was loaded into a coverage template feature class schema that included all of the attribute fields from the NTIA schema. The method of data loading was driven by the format in which it was received. Providers who supplied GIS shapefiles or feature classes could generally be loaded into the coverage template feature class schema using the simple data loader while CAD data had to be exported to GIS format prior to being loaded into the coverage template. Coverage data supplied as digital or analog maps required georectification and digitizing prior to loading into the coverage template feature class. Domain codes were entered in the appropriate attribute field for technology of transmission, maximum advertised downstream speed, maximum advertised upstream speed, and spectrum. If a provider did not identify the technology of transmission for any given coverage area, they were contacted by phone or email in order to obtain this information. When speed information was not included in the data submission, providers were contacted and asked to supply this data; in cases where the provider refused to give either the downstream, upstream, or both speeds, the lowest domain code was entered in the applicable attribute field. If a provider did not specify the type and spectrum used for fixed wireless the default values for unlicensed were used. Standardized confirmation maps were created for each provider by type of technology and sent to the provider for review. Once processing was completed for a providers census block submission, they were run through an ESRI geoprocessing model that performed several quality control-quality assurance tests and selected census blocks less than or equal to two square miles whose centroid was within the coverage area and road segments that intersected with census blocks greater than two square miles were clipped to the coverage area. Providers who submitted customer locations typically fell into four categories. Some were submitted as AutoCAD files where the points could be transferred to the GIS, then spatially joined to the census blocks they were located within. Others submitted maps in image format that were georectified in the same manner as other images, then census blocks were selected by an operator viewing the customer point images underlying the census blocks. When customer lists were submitted, they were loaded in a database and geocoded using ESRI Business Analyst USA Geocoding engine based on TeleAtlas road features. The geocoded points were subsequently treated identically to customer locations submitted in GIS or CAD format, and spatially joined to the census block template file.

OTHER LEVELS OF COARSE GEOGRAPHIC SUBMISSION:

This category had a wide range of submissions. The most common was as telephone exchange areas or equivalent, wire centers, zip codes, counties or general references to towns or cities. The problem with these submissions was that often a given polygon overlapped a census block or multiple blocks, and in most cases, they were much larger geographic entities than a census block.

HOW THEY WERE HANDLED:

Operational rules established early in the project did not allow provider coverage data that significantly over-represent provider coverage. Those providers that submitted coverage area by geographic features coarser than a census block that crossed county lines were not able to be processed. No interpolated data was used to calculate this data, if the data was not submitted by a provider in a format capable of processing the data was not calculated for that provider. Some providers who submitted broader geography initially that also were represented in the last point of aggregation infrastructure point file were sent estimated census block coverage maps and spreadsheets, and provided a second submission with finer level geography. Providers submitting town locations for DSL or Cable were handled differently, and used as validation for central offices from the last point of aggregation table, and subsequently to run the DSL modeling routine or validate a cable or cable plus areas. In instances where no infrastructure was identified with a reliable location (no verified street address or visual location), if it was a small town, typically smaller than 3 miles in width or length, then the DSL model was applied.

Final processing was the same regardless of the source process to derive provider coverage to the census block level. All technology of transmission types except fixed and mobile wireless was handled in a similar fashion. Outstanding questions remain to NTIA about fixed wireless, as to whether that technology fit in table 1(A) or in 1(B). No answer has been forthcoming, so we chose to include those in table 1(B). For DSL, Cable, Copper Wire and Fiber to the End User, the census block coverages were split into two categories, those less than 2 square miles, and those greater or equal to 2 square miles.

Those less than 2 square miles in area were left intact as census blocks. The census block coverage files greater than or equal to 2 square miles were intersected with the Tiger 2009 road files, splitting the Tiger road files at the census block boundary. A subsequent spatial join allowed the transfer of the data attributes in the census blocks greater than or equal to 2 square miles to be transferred to the Tiger road segments intersecting the census block coverage.

Based on comments made by NTIA in several of the webinar sessions, the preference was to use 2000 census topology and 2009 Tiger road files. The final NTIA NSGIC geodatabase that was recommended by NTIA as a preferred delivery format had parsed street attribute fields in the geodatabase schema. The Tiger 2000 roads had parsed street segment database fields for address, prefix, suffix, etc. The Tiger 2009 road data carried these as one field labeled "FULLNAME". The state does not plan to geocode using Tiger 2009 files, so it was not practical to do the large amount of manual work to parse the FULLNAME address field into individual components.

PROCESS DESCRIPTION
Solicit census block level broadband coverage areas from providers who initially submitted more generalized geographic coverage.

Several providers initially sent coverage at a broader level of geography than required, typically census tracts, telephone exchange areas, or zip codes. With public and commercial data on infrastructure, DSL and Cable models were prepared for each of these providers with best estimates of their coverage down to the finer granularity of a census block, along with the standardized spreadsheets with the details on each census block in the model. A dynamic web based map service was also made available to assist them in identifying census blocks and tracts with a Google map backdrop. Several providers subsequently used these tools and other analysis to submit more detailed coverage and data attributes in a second submittal.

PROCESS DESCRIPTION
Model Cable coverage (technology of transmission codes 40 & 41).

An ESRI geoprocessing model was created to generate coverage areas for Cable providers who did not submit census block or coverage data (i.e., census tract providers).

The most authoritative GIS layer available from the state with incorporated areas and city boundaries was used as a surrogate to model cable broadband coverage. Municipalities and towns were sporadic in their digital update of these maps, since annexations and other boundary modifications were ongoing and difficult to maintain in real time updates. To compensate, likely areas contiguous to these city boundaries were added, labeled "Cable-Plus" in the operational data model. These additional polygons were determined using operator interpretation, road density, structures points from Info USA in ESRI Business Analyst, and in some instances NAIP imagery. In general areas were added that were immediately contiguous to existing city or town boundaries that represented likely areas where cable service existed.

Cable broadband providers primarily work under the structure of franchise agreements with municipalities. Phone calls were made to the largest cities in the state in order to obtain that respective city's cable franchise agreement. They were all either unknown or a text agreement without maps.

The full set of potential cable areas were then passed through validation sources to determine if cable was provided. This included public sources, such as the Warren Communications Cable Fact book (http://www.warren-news.com/factbook.htm).

The second and most authoritative form of validation was data received from cable providers at the census tract, block, or coverage level of geography. A spatial join geoprocessing operation was performed on these datasets with the full set of potential cable coverage areas in order to further validate areas with cable coverage.

The third source of validation came from the public speed test site maintained throughout the project. Whenever user submitted speed tests identified cable modem broadband service near or adjacent to existing estimated cable areas, the cable-plus boundaries were expanded using the same method of digitizing outlined above.

It was not possible to differentiate between technology of transmission codes 40 and 41 using this indirect mapping method. The only authoritative way to determine this information was from data submitted by a provider. In all cases where the provider did not indicate the type of cable modem technology being used, the code for Cable Modem-Other (41) was assumed.

PROCESS DESCRIPTION

Model DSL coverage (technology of transmission codes 10 & 20).

An ESRI geoprocessing model was created to generate coverage areas for DSL providers who did not submit census block or coverage data (i.e., census tract providers). This model is based on typical DSL technology which can provide service up to 18,000 feet from a central office or remote terminal, unless otherwise specified by a provider. Since DSL lines are typically buried alongside roadways, underneath roadbeds, or strung on aerial telephone lines which tend to run alongside a road, a GIS dataset of a state’s road network could be used as a surrogate to model DSL areas. Commercial (GeoTel) and publicly available data sources representing last points of aggregation (LPA) for DSL were collected including central offices and remote terminals. Each LPA was validated based on publicly available data, provider data, and independent measurements; LPAs were used in a DSL model only if they were supplied directly from a provider or could be verified by two or more sources. The actual geoprocessing model used the validated central office and remote terminal locations to generate a raster cost surface based on all of the available roads radiating out 18,000 feet from each active LPA point. The raster coverage was converted to a polygon feature class and a small back-buffer was applied to achieve the final DSL coverage polygon representing a provider's maximum possible DSL coverage area. The DSL coverage areas were then used to select intersecting census blocks and

Remote terminals were provided or publicly available for only a small number of providers, therefore this method may tend to underestimate the full DSL coverage.

It was not possible to differentiate between ADSL or SDSL based on the LPA data; the only authoritative way to determine this was from data submitted by a provider. In all cases where the provider did not indicate which type of DSL service was being provided, the technology code was assigned to 10 "Asymmetric xDSL".

Process Description

Public broadband data research.

Provider presence maps were developed for central office locations and incumbent local exchange carrier locations for all assumed providers in the state. These were identified through a commercial spatial database purchased from GeoTel Inc., and supplemented by other public data sources such as the State's Public Service Commission and DSLReports.com. These were intended to be "talking maps" and general intelligence on where providers have infrastructure for subsequent phone and written communications with providers. These maps were compared to counties served by provider in the state's telecommunications association directory.

Web site research, review of materials submitted to the state by providers, and public websites, such as the FCC were researched for each provider.

Process Description

Solicit census block level broadband coverage areas from providers who initially submitted more generalized geographic coverage.

Several providers initially sent coverage at a broader level of geography than required, typically census tracts, telephone exchange areas, or zip codes. With public and commercial data on infrastructure, DSL and Cable models were prepared for each of these providers with best estimates of their coverage down to the finer granularity of a census block, along with the standardized spreadsheets with the details on each census block in the model. A dynamic web based map service was also made available to assist them in identifying census blocks and tracts with a Google map backdrop. Several providers subsequently used these tools and other analysis to submit more detailed coverage and data attributes in a second submittal.

Community Anchor Institutions (CAI)

Logical Consistency Report

All institutions on the initial draft spreadsheets used for the first two submittals were geocoded using ESRI Business Analyst Desktop with the USA Geocoding engine using TeleAtlas premium road features. This was judged to be the best available geocoding source for batch processing of addresses. No commercial source is 100% accurate in a primarily rural state such as this with low population numbers compared to other states and no large cities or metropolitan statistical areas. In every round of geocoding we used conservative matching criteria, and maintained and stored the type of match (building match, address match, or zip code match), along with a record of those not matching and not able to geocode.

COMPLETENESS REPORT

All geocoding is dependent on accurate road locations and complete and accurate street segment attribution. The GIS road layers available from the state were not judged as complete as the premium commercial sources. The Tiger 2009 road files, while spatially comparable to the commercial sources, have a large percentage of null values in the database attribution and street segment address ranges necessary for accurate geocoding. As in most parts of the country, geocoding is more accurate in urban settings than in rural routes. Complicating the process in a rural state for anchor institutions are the situation where some anchor institutions, such as public safety anchors are often staffed by volunteer staff and a post office box is the only valid address, and the physical address is wherever the public safety equipment is parked or stored at any given point in time.

PROCESS DESCRIPTION

Acquire lists of community anchor institutions.

Lists were obtained from the state and affiliated processional organizations for anchor institutions to be included in the broadband mapping in each of the community anchor institution community code categories. These were sorted and cross referenced and an initial round of elimination of duplication was accomplished.
PROCESS DATE 2010-06-15

PROCESS DESCRIPTION

Create and publish and process user speed test web site.

Created a public facing web site allowing anchor institutions to complete a brief survey and run a speed test on their connection using the Ookla speed test. The speed test site allowed a user to enter their location as an address on a simple Google map driven interface, and subsequently choose to move the location if it did not geocode to their satisfaction. Users were asked to select their technology of transmission from a list, enter their provider as a free text field, complete an optional questionnaire, and run a standard speed test on their connection. Behind the scenes, the date and time, and IP address of the user were captured.
All speed tests were geocoded, and the IP address was looked up in batch mode in the WHOIS database returning one or two providers registered with WHOIS. All speed tests registered between March 3, 2010 and February 14, 2011 were cleaned and analyzed against provider submissions and models.

A final provider assignment was assigned by examining the WHOIS fields, and the provider submitted by users.

There was considerable variation in the technology of transmission reported by users taking the speed tests. A final connection field was created and in most cases, the user selection was carried into this field. If a provider had only one confirmed technology of transmission, than all technologies listed by users were standardized to that, otherwise the user selection was carried. The state chose to not use the speed test data for an authoritative determination of the question if the institution subscribes to broadband service at the location. This was due to variability in user responses, the anonymity of the user submission and the lack of a practical mechanism for authoritative user identification. In future maintenance updates, the intent is to use the speed tests for anchor institutions in aggregate generalized analysis.

PROCESS DESCRIPTION

Geocode addresses and attribute anchor institutions.

All institutions on the initial draft spreadsheet were geocoded using ESRI Business Analyst Desktop with the USA Geocoding engine using TeleAtlas premium road features. This was judged to be the best available geocoding source for batch processing of addresses. No commercial source is 100% accurate in a primarily rural state such as this with low population numbers compared to other states and no large cities or metropolitan statistical areas. All geocoding is dependent on accurate road locations and complete and accurate street segment attribution. The GIS road layers available from the state were not judged as complete as the premium commercial sources. The Tiger 2009 road files, while spatially comparable to the commercial sources, have a large percentage of null values in the database attribution and street segment address ranges necessary for accurate geocoding. As in most parts of the country, geocoding is more accurate in urban settings than in rural routes. Complicating the process in a rural state for anchor institutions are the situation where some anchor institutions, such as public safety anchors are often staffed by volunteer staff and a post office box is the only valid address, and the physical address is wherever the public safety equipment is parked or stored at any given point in time.

PROCESS DESCRIPTION

Assign community anchor institution category codes.

Category codes were assigned based on the original source list and from keywords in the name of the institution and independent research. Technology of transmission and advertised speeds were obtained when possible, which initially was entirely based on the anchor institutions maintained by the state for consortiums providing state service contracts. Two iterations were accomplished with these state maintained lists, and all available attributes were obtained with assistance of the state analysts.

PROCESS DESCRIPTION

Cross reference anchor institutions against public data, research data attribution and delete duplicates.

After initial data collection, analysts worked on researching, calling and improving the addresses for those below an 80% match criteria. Many on the 70 percent matching range were fairly accurately located. The difference between a 70% and 80% match typically occurred when an address lacked a prefix or suffix cardinal direction on a street that had two cardinal directions (example 101 1st Street, on a street segment with 101 N. 1st Street and 101 S. 1st Street). Analysts were also able to obtain physical addresses for some lists supplied by the state with only a P.O. Box.

PROCESS DESCRIPTION
Final geocode addresses with corrections.

The lists with updated and corrected addresses were re-geocoded for the final mapping effort, and any anchor with any level of geocoding was included on the final map. The operational database identifies the type of match, so future maintenance cycles can be prioritized and targeted to those matching only zip codes or with address changes.
PROCESS DATE 2010-06-15

PROCESS DESCRIPTION
Overlay all remaining anchor institution points via spatial join on broadband coverages unioned and dissolved by concatenated provider/technology of transmission combinations.

Geocoded anchor institutions were spatially joined to unioned and dissolved concatenated provider/.technology of transmission combination broadband coverages. This provided some level of validation that an anchor at least was located within an area of available broadband coverage.

PROCESS DESCRIPTION
Populate technology of transmission, availability of broadband service, and maximum advertised download and upload speeds.

From the results of the previous step some attribution of database attributes for attributes with null values was accomplished. This step was rule based. The attribute of whether an anchor institution subscribes to broadband service could only authoritatively be answered yes, if the information was provided by the state, or a confirmation from an anchor speed test could be matched. Those anchors that were located within an area covered by a DSL, cable, other copper or fixed wireless were also assumed to have the ability to subscribe to broadband coverage and were also estimated to be subscribers. Assigning the technology of transmission and the advertised speeds (which required identifying a provider for the anchor institution) was only possible on a subset of all coverage in those areas where only one provider/technology of transmission was present. This allowed a few hundred more anchors to be identified, but typically only occurred in rural settings. Most urban settings had multiple providers. In addition

many providers submitted multiple technology options, so identifying one provider/technology of transmission combination was not possible even if there was only one provider possible for the anchor institution.
It is likely that in some instances in the rural settings and small towns an anchor institution may rely on mobile wireless broadband. This is common in public safety mobile equipment such as vehicles, but likely less common in anchor facilities. For the purpose of assigning attribution to anchor institutions with remaining null attributes, we took a conservative approach and did not overlay anchor institutions on mobile wireless coverages to assign attributes.

PROCESS DESCRIPTION

Maximum advertised downstream and upstream speeds were not available or collected for any of the CAIs.

A new domain value of "U" for Unknown was added to the data model for the current submission, and all values formerly coded as 0, were changed to "U".
PROCESS DATE 2011-03-01

PROCESS DESCRIPTION

A new optional attribute was requested by NTIA for the current submittal requesting knowledge about the presence or absence of WIFI at the CAI location.

This was not researched and attributed by the state in the current submission. All records were set to "Unknown" for the attribute, Public Wi-Fi.

PROCESS DESCRIPTION

In the first two submission processes for geocoding we used conservative matching criteria, and maintained and stored the type of match (building match, address match, or zip code match), along with a record of those not matching and not able to geocode. The current submission was completed by state analysts, and new additions to the list were not geocoded. The additions of new anchor institutions in this submission were assigned their latitude and longitude geographic location based on their location used in the Montana Structures Framework.

PROCESS DESCRIPTION

A new optional attribute was requested by NTIA for the current submittal requesting a CAI unique identification number for K-12 schools, libraries and colleges and universities. The following steps were completed for this request.

1. Added CAIID for the Library category using a combination of the FSCSKEY and FSCS_SEQ number attributes from http://harvester.census.gov/imls/data/pls/index.asp. Added 49 records using the Montana Structures Framework to assign their geographic location.

2. Added CAIID for the University, college, other post-secondary category using the NCES IPEDS ID from http://nces.ed.gov/ipeds/datacenter/. Added 10 records using the Montana Structures Framework to assign their geographic location.
3. Added CAIID for the School – K through 12 category using the NCES ID CCD ID from http://nces.ed.gov/ccd/bat/ . Added 118 schools using information from the OPI Schools http://www.publiclibraries.com/montana.htm list, the NCES Schools List and the Montana Structures Framework. NOTE: NTIA asked that each school be given a unique ID but in the CAI table, many schools at the same address were combined. These were not separated for this round of the NTIA submittal.

PROCESS DESCRIPTION

A new optional attribute was requested by NTIA for the current submittal requesting a URL for each anchor institution.

Assigned URLS to CAI records: for the University, college, other post-secondary category assigned the actual URL for that institution; for the Library category added a standard URL (http://www.publiclibraries.com/montana.htm); for the School – K through 12 category added the OPI URL (http://opi.mt.gov/Resources/Directory/Index.html); and for other institutions, added an appropriate URL for the type of CAI.

WIRELESS SERVICE COVERAGE (BB_Service_Wireless)

LOGICAL CONSISTENCY REPORT

Data submitted by broadband providers was accepted as is when it was provided as a broadband coverage or at a census block level. Provider coverage submitted at a coarser geographic scale was supplemented with public data, independent measurements and GIS modeling techniques. When independent measurements were available for typical broadband speeds and modeled location of infrastructure, some provider data was overridden or supplemented.

COMPLETENESS REPORT

All data submitted by broadband providers was mapped in complete form, except where independent measurements were used to supplement provider submittals. Several providers did not participate in the broadband mapping process, including some that were suspected to be providers.

PROCESS DESCRIPTION

Public broadband data research.

Two forms of wireless coverage were provided in this table, fixed point to point wireless and mobile wireless. Outstanding questions remain to NTIA about fixed wireless, as to whether that

technology fit in table 1(A) or in 1(B). No answer has been forthcoming, so we chose to apply those in table 1(B). No public data was located on fixed wireless infrastructure points, except notification of availability on provider's web pages, and in some instances, specific towns, recreation or commercial locations where wireless service was provided. No modeling was attempted on fixed wireless coverage. All coverage came directly from providers or was mapped from locations provided on a provider web page. We did not attempt any propagation modeling on fixed wireless, since that can be influenced by local structures and vegetation in the vicinity. A few providers did provide coverages that appeared to be derived from propagation modeling.
Most of the public data research focused on mobile wireless providers using cellular service spectrums. The Federal Communications Commission (FCC) Universal Licensing System (ULS) is the consolidated database and application filing system for most Wireless Radio Services. ULS supports electronic filing and provides public access to licensing information, weekly Public Notices, FCC rulemakings, processing utilities, a telecommunications glossary, and much more." The FCC ULS Advanced Licensing Search was queried for all FCC licenses filed in the state; a relational database was built from the results. Information from the database was extracted in order to perform the cellular tower propagation modeling for wireless broadband.
The FCC ALS and ULS reporting systems were the source for most of the tower locations. Towers were required to be licensed when they meet specific published criteria. These included some variables that could be modeled with GIS statewide, such as varying proximity to airports and heliports, combined with specific local level criteria not easily obtained or modeled statewide such as the grade construction within proximity of these, and any structure over 200 ft in height. A number of cell towers providing broadband were likely not located in the FCC database. None of the mobile wireless providers were willing to provide infrastructure such as tower locations and parameters, and the coverages provided were very generalized.
The mobile wireless coverage in the state is in transition. There were currently no GSM mobile wireless providers meeting the NOFA criteria for being a provider. There is some GSM infrastructure in the state maintained for wholesale arrangements and roaming users with GSM technology.

PROCESS DESCRIPTION

Create and publish user speed test web site.

A public facing website was created in the spring of 2010 asking internet users in the state to complete a brief survey regarding their internet connection and run a speed test on their connection using the Ookla speed test. The speed test site asked that a user enter their location as an address on a Google map interface. If the address did not geocode to their satisfaction, the user could choose to move the place mark to their desired location. Next, users were asked to select their technology of transmission from a list, enter their provider in a free form text field, complete an optional questionnaire, and run a standard speed test on their connection. The date and time, and IP address of the user were captured during the speed test.

All speed tests were geocoded, and the IP address was looked up in batch mode in the WHOIS database returning one or two providers registered with WHOIS. All speed tests registered between March 3, 2010 and May 14, 2010 were cleaned and analyzed against provider submissions and models for the first and second data submissions. For the current (third) submission, all the data between March 3, 2010 and February 15, 2011 was standardized and used.

For the first two submissions a final provider assignment was assigned by examining the WHOIS fields, and the provider submitted by users. Consistent rules were not always possible, but generally when two WHOIS records were returned, the second more specific WHOIS provider was selected. In some instances, where the WHOIS providers were backhaul or other and were not providers meeting the NOFA criteria, the user submitted provider designation was cleaned and standardized and assigned as the final provider. For the current submission a rule based database program was built by the Montana Dept. of Administration to automate the final provider assignment.

There was considerable variation between the user reported technology of transmission (TOT) and the known technologies for any given provider. Records were divided on unique provider/ TOT combinations for the first and second submissions, which limited the record count in many instances. For the current submission the records were divided only by provider, not taking TOT into consideration.

For the first two submissions, the speed test records were used in two ways for the final processing.

1) As an independent measurement to validate the presence/absence of a provider coverage for DSL and/or Cable technologies.
In the first submission a few providers were identified as DSL broadband providers based primarily on speed tests. In these instances, DSL models were executed for both providers based on verified central office locations. Some Speed tests with an identified technology of transmission of Cable Modem were used to expand “likely” cable areas which were typically adjacent to incorporated and urban areas. These “cable-plus” areas were created to supplement submissions from Cable Modem providers who did not provide detailed coverage or census blocks. No new DSL providers or Cable providers were identified using speed tests in the current submission.

2) As an independent measurement for typical upload and download speeds.
Once data were cleaned and final provider and technology of transmission assigned, these fields were concatenated. In the first two submissions, if the remaining records exceeded 10 for the combination of provider and technology, and the speed test was successfully completed (values > 0) the average value and standard deviation of the download speed were calculated. Any values exceeding 1 standard deviation were removed as outliers, and the mean of the remaining records within 1 standard deviation was calculated for the download and upload speed. This value was reported for each provider/technology of transmission record as the typical speeds for that provider. In some instances the typical speed was lower than that required to meet the definition of broadband by NTIA, but that did not preclude the records from being included in the broadband map in the first two submissions as it did in the current submission.

For the current submission, these procedures were modified and all records were re-run. The steps of the current processing are provided below. The primary procedural change was to drop the validation of the presence/absence of provider coverage for DSL and/or Cable technologies, since providers had been validated in the first two submissions and potential new providers identified through additional speed tests were determined to not meet the NOFA criteria for being considered a broadband provider. The use of the speed test data for determining typical speeds was implemented with similar rules as the first two submissions with the exception of the use of the technology of transfer, and raising the minimum number of speed tests to 15, after removing outliers, to be used in typical speed calculations. Procedurally, the process was also automated with a Python script to improve processing performance and minimize quality control/quality assurance testing.

Typical upload and download speeds for all providers with less than 15 processed speed test records were coded as null values. In addition, based on telephone communication with NTIA on March 9, 2011, all typical speeds less than minimum NOFA upload of download speed criteria were also ignored and reported as null. Based on a related request in the same communication, the typical speeds greater than the advertised speeds were ignored and reported as null. Subsequently on March 17, in the NTIA grantee webinar, the NTIA staff indicated that typical speeds would not be compared to advertised speeds. This rule change was not received in time to implement in the workflow for the current submission, and will be implemented in the fourth submission in the fall, 2011. We anticipate other significant modifications in the use of the speed test data for the next submission, since many of the records older than one year will not be used in future calculations.

Processing steps for the current submission are provided below:

1. Speed test records were imported into a SQL Server data file, adding fields Final Provider and IPGroup to the initial records.
2. IPGroup attribute was set by extracting the left three nodes of the IP Address of the speed test (e.g. 161.7.1.236 had 161.7.1) moved to the IPGroup attribute.
3. An IPGroup to Final Provider cross reference table was created to determine the final provider from the unique three part IPGroup (e.g. 161.7.1 is known to be the State of Montana).
4. Each IPGroup was reviewed with the data in the who is 1 provider, who is 2 provider and then the user specified provider to determine the most authoritative final provider from the official list of providers. None of the WHOIS or user submitted fields were absolutely authoritative in all instances, so expert opinion by technicians knowledgable of the providers was used in some instances to assign the IPGroups, and subsequently the Final Provider attribute.
5. Run a python script to remove outliers and calculate summary statistics for each Final Provider assignment. The rationale for removing outliers was to mitigate the many variables that effect a typical speeed test, such as the time of day, others on the network, etc. The script implemented the following work flow rules:

a. Use all records for each unique FinalProv attribute value with D_kbps greater than 0 or U_kbps greater than 0 , then:
b. Calculate a mean for the unique provider group for each D_kbps and U_kbps.

c. Calculate a standard deviation for the unique provider group for each D_kbps and U_kbps. Each speed attribute was calculated independently of the other.
d. Subtract the outliers (if any) higher or lower than one standard deviation from the mean.
e. Calculate the median value of the remaining non-outliers for each provider D_kbps and U_kbps respectively.
f. Create a summary table with the final calculated assignment of FinalProv, D_kbps and U_kbps.

6. Post process the summary table in the following sub steps:

a. Join the summary tables by provider for the upload and download speeds into one summary file including the number of records or frequencies for up and down speeds for each provider after removing the outliers, and the mean up and down speeds in kilobits per second for each provider.
b. Select "FreqDown" < 15 AND "FreqUp" < 15 then delete the resulting selection set from the joined table. The FreqDown/Up fields counted the number of speed test records for a provider after the outliers more or less than one standard deviation from the mean value were removed from consideration.
c. Select "D2_kbps" <= 768 kbps AND "U2_kbps" <= 200 kbps. then delete the resulting selection set from the joined table.

7. Import the remaining valid mean values for each provider into the appropriate broadband coverage feature classes.
8. Select any typical speeds greater than advertised speeds either up or down, and make the resulting records null in the final broadband coverage feature classes (as per NTIA request 3/9/2011).

PROCESS DESCRIPTION

Creation, processing and maintenance of last point of aggregation for infrastructure.

Any fixed or mobile wireless antenna or tower location submitted by a provider, or obtained from the FCC that was used in the final processing for wireless broadband coverage was maintained in the operational database for last point of aggregation, and subsequently transferred to Table 3 backhaul and middle mile points.

PROCESS DESCRIPTION

Solicit broadband coverage areas, infrastructure, advertised and typical speeds and components of subscriber weighted nominal speed from providers.

Requests were made via email and phone calls to every broadband provider to provide coverage, technology of transmission, advertised and typical download and upload speed at the census block level.

All data types were accepted as submittals and a large variety of data were submitted, including narrative descriptions, coverage maps as JPG or PDF images, CAD files, GIS shapefiles, KMZ and KML files, FCC form 477 reports , and data spreadsheets.

Providers that submitted actual broadband coverage data or census block representation of coverage along with the components necessary to complete the data attribution were not contacted for a follow up. The providers who did not submit data based on initial queries were sent a standardized spreadsheet to map their coverage and optionally infrastructure, and subsequently a final notice email.

PROCESS DESCRIPTION

Fixed Wireless.

Providers submitted coverage data in a wide variety of formats, levels of completeness, and at varying geographic scales. All types of data was accommodated and processed whenever possible. An open structure process for submittals was allowed, accepting any data, and attempting to work with the provider when questions arose. If data was submitted by a provider in a format that did not allow a direct coverage to be mapped, such as a coarse level of geography such as a census tract, or county, feedback was provided to the providers in the form of standardized spreadsheets in an attempt to standardize the inputs, and increase the geographic granularity of the provider data submission. Although each provider had individual characteristics and nuances in their data submissions, some data patterns can be described generalizing the typical types of submissions. In general, for fixed wireless to be mapped it was necessary to receive data from a provider, since there were no public sources available on point to point wireless tower locations in public form, except as depicted on providers web pages in a few instances.

FCC FORM 477 REPORT OR SIMILAR FORMAT:
Geographically, these were lists of census tracts of coverage, accompanied by additional documentation on technology of transmission, speed tiers, and number of customers. Providers submit these twice a year to the FCC and recent submissions have been done using a structured web site maintained by the FCC. A few providers submitted printouts that appeared to be from this web format and were typically complete and standardized. More providers submitted spreadsheets roughly in the F477 format, but with modified and generalized data.

HOW THEY WERE HANDLED:
If the providers identified specific coverage areas as census blocks, or direct coverage area, or as infrastructure tower locations, they were processed and mapped. Providers identifying census blocks were processed by dissolving the census blocks into single coverage polygons by speed tier. Providers identifying a direct coverage area were converted directly to GIS polygon files and attributed. Providers submitting tower locations were mapped as circular polygons centered on the tower with a radius averaging 10 miles measured as Euclidian (straight line) distance from the tower. Providers that specified variable radius were mapped as circles at the radius they submitted.

PROVIDERS SUBMITTING CENSUS BLOCK COVERAGE:
A few providers submitted coverage as census blocks, either through a tabular listing of census blocks or spreadsheet, or in map format. It was common that a provider where public data

indicated multiple technologies of transmission only submitted some of the technologies of transmission.

HOW THEY WERE HANDLED:
These were loaded directly into the master Census 2000 block coverage by provider and attributed with available data submitted by the provider. In instances where some data attributes were missing, such as advertised or typical speed tiers, or subscriber data, the data attributes were left blank or null. Providers identifying census blocks were processed by dissolving the census blocks into single coverage polygons by speed tier. A visual inspection of independent speed test data overlaying the provider submitted block coverage was completed, but no action was taken to override a provider's submittal.

PROVIDERS SUBMITTING ACTUAL COVERAGE MAPS:
Coverage maps were submitted by several providers, or coverages were derived from public sources or from other indirect indicators of coverage such as customer point maps or tabular lists in text or spreadsheet format. These were differentiated from the other types of geographic submission coarser than a census block since they represented the full and explicit range of coverage.

HOW THEY WERE HANDLED:
Coverage maps were treated as explicit coverage and all census blocks intersecting any portion of a coverage were selected and attributed with the provider coverage by technology of transmission, and all related attributes were transferred to the census block representation. The method of creating the coverage varied by source. Providers who supplied broadband coverage as a GIS polygon or CAD feature were converted to polygons. Some providers, including non-responsive providers who did not submit anything to the project, had published coverage maps of various forms on their web sites or submitted an image in jpg, tiff, pdf or other graphic format. These were geogeorectified to base map layers, typically roads, but sometimes other features such as state or county boundaries or towns, and subsequently converted to polygon features. Then they were intersected and transferred to census block feature classes like the digital GIS submissions. Providers who submitted customer locations typically fell into four categories. Some were submitted as AutoCAD files where the points could be transferred to the GIS, then spatially joined to the census blocks they were located within. Others submitted maps in image format that were georectified in the same manner as other images, then census blocks were selected by an operator viewing the customer point images underlying the census blocks. When customer lists were submitted, they were loaded in a database and geocoded using ESRI Business Analyst USA Geocoding engine based on TeleAtlas road features. The geocoded points were subsequently treated identically to customer locations submitted in GIS or CAD format, and spatially joined to the census block template file. A visual inspection of independent speed test data overlaying the provider submitted block coverage was completed, but no action was taken to override a provider's submittal.

OTHER LEVELS OF COARSE GEOGRAPHIC SUBMISSION:
This category had a wide range of submissions. The most common was as telephone exchange areas or equivalent, wire centers, zip codes, counties or general references to towns or cities. The

problem with these submissions was that often a given polygon overlapped a census block or multiple blocks, and in most cases, they were much larger geographic entities than a census block.

HOW THEY WERE HANDLED:
Our operating rules established early in the project did not allow final provider coverage to significantly over represent provider coverage. Those providers that submitted coverage area by coarse geographic features and did not specifically identify coverage as a coverage layer or census blocks were not able to be processed. No interpolated data was used to calculate these data, if the data was not provided by a provider in a format capable of processing; the data was not calculated for that provider.

PROCESS DESCRIPTION

Mobile Wireless Verizon.

Where cell tower locations for specific providers could be consistently identified, propagation models of wireless coverage were developed for mobile wireless. SoftWright's Terrain Analysis Package software with the Longley-Rice algorithm for model development was used to develop the models. The models were constrained in a 25 radius of the tower. Output grid size was .5 kilometers. Areas that had coverage with signal strength above 40 dbu was classified as having broadband availability. All propagation data meeting the above criteria were merged into a single geodatabase. Non- contiguous areas of less than .5 kilometers were removed from the coverage to climate scatter that was deemed to be an artifact of data processing limitations. Verizon Wireless was the only provider where this method was possible to apply. A mobile wireless coverage was not provided by Verizon for the June, 2010 submission. In August, 2010 Verizon provided a coverage that used an estimated one-quarter mile raster resolution, and the digital elevation model TetraTech used for this analysis and propagation modeling was based on 30 meter resolution. As a result individual TetraTech's coverage for Verizon shows a slightly more spotty pattern where influenced by local topographic characteristics. Due to some missing or unidentifiable towers, we anticipate that coverage may exist in some areas that were not indicated as coverage. Feedback from Verizon is needed to resolve differences, and is scheduled for the 2011 updates. As a result, in the September, 2010 update we submitted the actual coverage provided by Verizon, replacing the TetraTech model submitted in June. Provider submittals did not differentiate download or upload speeds by tower location, only statewide. The typical upload and download speeds for Verizon were reported based on the mean upload and download point samples from several hundred thousand speed test samples, averaged statewide. Since tower identification numbers are rotated by providers, it was not possible, without provider data input, to validate 3G speed availability on a tower by tower basis. CDMA technology does allow a switch from 2G to 3G via software, so it is more likely that most if not all CDMA towers in the state can theoretically provide broadband coverage. An adequate number of speed tests were received for Verizon wireless, but the RootWireless data was more robust and hundreds of thousands of sample points were available, so the public speed test submittals were not used for this provider. All mobile wireless coverage were processed as single coverage files. All polygons had the same data attributes carried in all polygons. For the third submission, April 2011 Verizon submitted a new and slightly revised wireless coverage map that was used for the submittal.

PROCESS DESCRIPTION

Mobile Wireless Alltel.

Alltel did not provide a mobile wireless coverage in any form for Montana. We did not have independent measurements for this provider that could reliably be used. Rootwireless drove several thousand miles on state and federal highways in the state and provided ongoing point samples from several handsets with speed tests every 2 min and signal strength every 10 seconds during operation. Provider submittals did not differentiate download or upload speeds by tower location, only statewide. The typical upload and download speeds for Verizon were reported based on the mean upload and download point samples from several hundred thousand speed test samples, averaged statewide. Since tower identification numbers are rotated by providers, it was not possible, without provider data input, to validate 3G speed availability on a tower by tower basis. CDMA technology does allow a switch from 2G to 3G via software, so it is more likely that most if not all CDMA towers in the state can theoretically provide broadband coverage. For the first and second and third submissions, June and September, 2010 and April, 2011, a coarse, generalized map available in Alltel retail outlets and on their web site was used as a base for their mobile wireless coverage. The image was georectified and used in heads up digitizing to select 6th code watersheds obtained from the Montana State Library, that matched the Alltel coarse map coverage as closely as possible. With the coarse nature of the small scale Alltel map, we determined that an established map layer with appropriate sized polygons that could be mapped in a consistent and repeatable manner would be the best way to depict Alltel mobile wireless coverage. The minimum number of speed tests from the public speed test site was not achieved for mobile wireless to allow independent typical speed measurements for Alltel, so those remain null values in Montana. All mobile wireless coverage were processed as single coverage files. All polygons in the two mobile provider coverage in the state had the same data attributes carried in all polygons.

PROCESS DESCRIPTION

Satellite

The parameters below show the satellite wireless models for MT. A few satellites are use the same azimuth and altitude, so they only need to be run once and subsequently copied and renamed for different providers. There was one coverage for WildBlue and Starband, and four coverage for Hughes/DirectTV. The Anik-F2 satellite appears to be shared by Hughes and WildBlue coverage, and was listed under both.

The process included running a hillshade with the parameters shown below, selecting the "Model shadows" parameter. This was reclassed into 3 classes 0,1,Max value. Then the Majority filter model in Spatial Analyst Generalization was run with a 4x4 neighborhood grid to filter out the smallest isolated shadow pixels. A conditional selection of the class 0 (shadow values) was made for the final grid. This was run through a raster to polygon conversion and added to the master coverage template from geodatabase.

Provider	Satellite	Azimuth	Altitude	Operator
Hughes / DirectTV				
	Anik-E2	141.6	33.7	Telesat Canada Ltd.
	Anik-F2	181.8	36.13	Telesat Canada Ltd.
	Spaceway-1	170.6	35.68	Direct TV, Inc.
	Spaceway-3	160.1	34.17	Hughes Network Systems
WildBlue				
	Anik-F2	181.8	36.13	Telesat Canada Ltd.
	Wildblue 1	181.8	36.1	Wildblue Communications
Starband				
	Echostar 9	195.1	35.03	Echostar Technologies, LLC

Process Description

Quality assurance testing. A separate analyst checked each provider submission. Due to the variety of provider submissions, the analyst originally doing the work and the analyst checking discussed the interpretations when the criteria were subject to interpretation. Coverage, technology of transmission, and speed tier were checked completely for each provider.

Speedtest Data Processing

Process Description

Create and publish and process user speed test web site.

A public facing website was created in the spring of 2010 asking internet users in the state to complete a brief survey regarding their internet connection and run a speed test on their connection using the Ookla speed test. The speed test site asked that a user enter their location as an address on a Google map interface. If the address did not geocode to their satisfaction, the user could choose to move the place mark to their desired location. Next, users were asked to select their technology of transmission from a list, enter their provider in a free form text field, complete an optional questionnaire, and run a standard speed test on their connection. The date and time, and IP address of the user were captured during the speed test.

All speed tests were geocoded, and the IP address was looked up in batch mode in the WHOIS database returning one or two providers registered with WHOIS. All speed tests registered between March 3, 2010 and May 14, 2010 were cleaned and analyzed against provider submissions and models for the first and second data submissions. For the current (third) submission, all the data between March 3, 2010 and February 15, 2011 was standardized and used.

For the first two submissions a final provider assignment was assigned by examining the WHOIS fields, and the provider submitted by users. Consistent rules were not always possible, but generally when two WHOIS records were returned, the second more specific WHOIS provider was selected. In some instances, where the WHOIS providers were backhaul or other and were not providers meeting the NOFA criteria, the user submitted provider designation was cleaned and standardized and assigned as the final provider. For the current submission a rule based database program was built by the Montana Dept. of Administration to automate the final provider assignment.

There was considerable variation between the user reported technology of transmission (TOT) and the known technologies for any given provider. Records were divided on unique provider/ TOT combinations for the first and second submissions, which limited the record count in many instances. For the current submission the records were divided only by provider, not taking TOT into consideration.

For the first two submissions, the speed test records were used in two ways for the final processing.

1) As an independent measurement to validate the presence/absence of a provider coverage for DSL and/or Cable technologies.

In the first submission a few providers were identified as DSL broadband providers based primarily on speed tests. In these instances, DSL models were executed for both providers based on verified central office locations. Some Speed tests with an identified technology of transmission of Cable Modem were used to expand “likely” cable areas which were typically adjacent to incorporated and urban areas. These “cable-plus” areas were created to supplement submissions from Cable Modem providers who did not provide detailed coverage or census blocks. No new DSL providers or Cable providers were identified using speed tests in the current submission.

2) As an independent measurement for typical upload and download speeds.

Once data were cleaned and final provider and technology of transmission assigned, these fields were concatenated. In the first two submissions, if the remaining records exceeded 10 for the combination of provider and technology, and the speed test was successfully completed (values > 0) the average value and standard deviation of the download speed were calculated. Any values exceeding 1 standard deviation were removed as outliers, and the mean of the remaining records within 1 standard deviation was calculated for the download and upload speed. This value was reported for each provider/technology of transmission record as the typical speeds for that provider. In some instances the typical speed was lower than that required to meet the definition of broadband by NTIA, but that did not preclude the records from being included in the broadband map in the first two submissions as it did in the current submission.

For the current submission, these procedures were modified and all records were re-run. The steps of the current processing are provided below. The primary procedural change was to drop the validation of the presence/absence of provider coverage for DSL and/or Cable technologies, since providers had been validated in the first two submissions and potential new providers identified through additional speed tests were determined to not meet the NOFA criteria for being considered a broadband provider. The use of the speed test data for determining typical speeds was implemented with similar rules as the first two submissions with the exception of the use of the technology of transfer, and raising the minimum number of speed tests to 15, after removing outliers, to be used in typical speed calculations. Procedurally, the process was also automated with a Python script to improve processing performance and minimize quality control/quality assurance testing.

Typical upload and download speeds for all providers with less than 15 processed speed test records were coded as null values. In addition, based on telephone communication with NTIA on March 9, 2011, all typical speeds less than minimum NOFA upload of download speed criteria were also ignored and reported as null. Based on a related request in the same communication, the typical speeds greater than the advertised speeds were ignored and reported as null. Subsequently on March 17, in the NTIA grantee webinar, the NTIA staff indicated that typical speeds would not be compared to advertised speeds. This rule change was not received in time to implement in the workflow for the current submission, and will be implemented in the fourth submission in the fall, 2011. We anticipate other significant modifications in the use of the speed test data for the next submission, since many of the records older than one year will not be used in future calculations.

Processing steps for the current submission are provided below:

1. Speed test records were imported into a SQL Server data file, adding fields Final Provider and IPGroup to the initial records.

2. IPGroup attribute was set by extracting the left three nodes of the IP Address of the speed test (e.g. 161.7.1.236 had 161.7.1) moved to the IPGroup attribute.

3. An IPGroup to Final Provider cross reference table was created to determine the final provider from the unique three part IPGroup (e.g. 161.7.1 is known to be the State of Montana).

4. Each IPGroup was reviewed with the data in the who is 1 provider, who is 2 provider and then the user specified provider to determine the most authoritative final provider from the official list of providers. None of the WHOIS or user submitted fields were absolutely authoritative in all instances, so expert opinion by technicians knowledgable of the providers was used in some instances to assign the IPGroups, and subsequently the Final Provider attribute.

5. Run a python script to remove outliers and calculate summary statistics for each Final Provider assignment. The rationale for removing outliers was to mitigate the many variables that effect a typical speeed test, such as the time of day, others on the network, etc. The script implemented the following work flow rules:

   a. Use all records for each unique FinalProv attribute value with D_kbps greater than 0 or U_kbps greater than 0 , then:

   b. Calculate a mean for the unique provider group for each D_kbps and U_kbps.

   c. Calculate a standard deviation for the unique provider group for each D_kbps and U_kbps. Each speed attribute was calculated independently of the other.

   d. Subtract the outliers (if any) higher or lower than one standard deviation from the mean.

   e. Calculate the median value of the remaining non-outliers for each provider D_kbps and U_kbps respectively.

   f. Create a summary table with the final calculated assignment of FinalProv, D_kbps and U_kbps.

6. Post process the summary table in the following sub steps:

   a. Join the summary tables by provider for the upload and download speeds into one summary file including the number of records or frequencies for up and down speeds for each provider after removing the outliers, and the mean up and down speeds in kilobits per second for each provider.

   b. Select "FreqDown" < 15 AND "FreqUp" < 15 then delete the resulting selection set from the joined table. The FreqDown/Up fields counted the number of speed test records for a provider after the outliers more or less than one standard deviation from the mean value were removed from consideration.

   c. Select "D2_kbps" <= 768 kbps AND "U2_kbps" <= 200 kbps. then delete the resulting selection set from the joined table.

7. Import the remaining valid mean values for each provider into the appropriate broadband coverage feature classes.

8. Select any typical speeds greater than advertised speeds either up or down, and make the resulting records null in the final broadband coverage feature classes (as per NTIA request 3/9/2011).

Quality Assurance Testing

A separate analyst checked each provider submission. Due to the variety of provider submissions, the analyst originally doing the work and the analyst checking discussed the interpretations when the criteria were subject to interpretation.

Coverage, technology of transmission, and speed tier were checked completely for each provider.

Many of the models and block, tract and coverage level processes were completed with ESRI Modelbuilder and Python scripts, and these methods were tested for quality assurance in the preliminary mapping stages and in the initial sample data submissions to NTIA.

All providers who submitted geographic coverage coarser than a census block were provided a data checking package to assess for accuracy and completeness. Any comments received from providers were processed.

1. QA/QC Checks prior to Individual Data Processing (i.e., block or coverage geoprocessing model). [Automated Modelbuilder tools and follow-up by an analyst]
    a. Check for inconsistencies within the Provider Name, DBA Name, FRN
    b. Check for duplicate census blocks or coverage areas
    c. Check the Provider Name, DBA Name, FRN against the "Official Provider Table"

2. For each provider after initial data processing is completed [Review by an analyst that did not process the original data]
    a. Review correspondence log
        i. Review recent correspondence, since previous NTIA submission
        ii. Note changes/additions/comments on coverage area, technologies, speeds, infrastructure, subscriber weighted nominal speeds (SWNS)
    b. Review wiki data processing page (current metadata)
        i. Note changes/additions/comments on coverage area, technologies, speeds, infrastructure, SWNS
    c. Review individual Provider Wiki page (historic metadata)
        i. Note changes/additions/comments on coverage area, technologies, speeds, infrastructure, SWNS
    d. Check Provider Data Folder 2011_03
        i. Review recent data submissions, since previous NTIA submission
    e. Check Working Data Folder 2011_03
        i. Review current update feature class geography
        ii. Review coverage with provider's submissions
        iii. Review technology of transmissions (TOTs) with provider's submissions
        iv. Review Max Adv Speeds: Down/Up with provider's submissions

3. For each provider after final data processing is completed [Review by an analyst that did not process the original data]
    a. Check PROVCOV_Master geodatabase:Provider Blocks feature class and/or Provider Coverage feature class
        i. Review geography
        ii. Review TOTS
        iii. Review Max Adv Speeds: Down/Up

4. Check Infrastructure feature class [Review by an analyst that did not process the original data]
    a. Review recent submissions, since previous NTIA submission

5. Check SWNS feature class [Review by an analyst that did not process the original data]
    a. Determine if provider submission is valid

6. For each provider after speed tests are processed [Review by an analyst that did not process the original data]
    a. Check PROVCOV_Master geodatabase for Typical Speeds: Down/Up

7. QA/QC Checks and Reports on the Final NTIA Deliverable [Automated Modelbuilder tools and follow-up by an analyst]
    a. Check the Provider Name, DBA Name, FRN against the "Official Provider Table" for each NTIA feature class (i.e., BB_Service_CensusBlock, BB_Service_RoadSegment, BB_Service_Wireless, etc.). NTIA_Provider_Name_DBA_FRN_Errors_Sample.xls, looks at each NTIA feature class (i.e., census blocks, road segments, wireless, etc…) and checks to see if there is an identical match in the "Official Provider Table." If an identical match does not exist for that Provider Name, DBA Name, FRN concatenation it is written to a geodatabase table along with the NTIA feature class where the "error" occurred. When an "error" does occur it then has to be checked by an analyst and corrected if necessary.
    b. Change Detection Report – This geoprocessing model compares and reports any changes in the Census Block, Road Segment, and Wireless feature classes for the current and previous versions of the NTIA SBDD Transfer database. The user needs to supply the feature classes for each NTIA version as well as the name of the final change detection table. NTIA_Change_Detection_Example.xls, compares and reports any changes (limited to Provider Name, DBA Name, FRN, TOT combinations) in the Census Block, Road Segment, and Wireless feature classes for the current and previous versions of the NTIA SBDD Transfer database. If the final change detection table has no records, then no changes were detected between the two databases. If a Provider Name, DBA Name, FRN, TOT combination does not have a "pair" in either direction (the current or previous

NTIA database) then it is written to a geodatabase table along with the NTIA feature class and version where the "error" occurred. This report does not change any data in either database but rather acts as a flag, requiring an analyst to check if the "error" is valid.

c. Check for duplicate census blocks or road segments or wireless coverage areas.
d. Check for duplicate anchor institution points.

8. Review Final NTIA deliverables [Review by an analyst that did not process the original data]
    a. Review BB_ConnectionPoint_MiddleMile
    b. Review BB_Service_CAInstitutions
    c. Review BB_Service_Census Block
    d. Review BB_Service_Overview
    e. Review BB_Service_RoadSegment
    f. Review BB_Service_Wireless

9. Run the NTIA Check submission tool and python tool to confirm that all possible records passed the NTIA data checks. The only items that failed in the checking process were those where inconsistencies in the final NTIA NSGIC data model did not agree with the final documentation and rules established by NTIA and FCC in the final webinar and documentation presented March 17, 2011. These exceptions were documented along with the submission.