State Broadband Data and Development Mapping Methodology

For the States of Alabama, Idaho, Wisconsin and Wyoming
Revised September 30, 2010
Redacted January 21, 2011

CostQuest Associates

LinkAMERICA Alliance

Contents

Overview

The following documentation provides an overview of how the second required data set was collected and processed for the State Broadband Data and Development Program (SBDDP) in the states of Alabama, Idaho, Wisconsin, and Wyoming. As expected, this document rests heavily on the prior draft, but has also been updated and expanded.

Significant changes include additions covering:

1. User responses and lessons learned from online maps
2. Impact of program change and NTIA guidance
3. Confidentiality and the use of licensed materials

Treatment of the following subjects has been expanded:

1. Community anchor institutions
2. Verification and validation
3. Data production methods

As anticipated, the SBDD program continues to mature and evolve. Technical leadership and strong guidance has been appreciated. The opportunity to look at a three to five year project window has also let the team move forward in establishing more long term business processes. Our sincere hope is that these trends continue.

We close this document with an Appendix (Appendix One- Data Collection Challenges). This section describes some of the open issues, challenges and questions we are exploring. Our hope is to receive clarification and counsel from NTIA in how best to confront some of these issues which are likely common across states.

In our view, the second mapping deliverable reflects (1) a good faith effort, which results in a reasoned response to the NOFA, Technical Appendix A, as well as supplementary program office guidance offered in phone calls, emails, and webinars, (2) a stable foundation for improvement and prioritization of both NTIA and state needs and interests , (3) a valid data model to support online mapping, consumer feedback, provider verification and reporting, and finally, (4) a valid use of the evolving data transfer model and its intrinsic validation methods. More importantly, the resulting data and online coverage maps that follow from this work are providing good input and context for the broadband planning teams working across the states we have the pleasure to serve.

Purpose of This Manual

This technical document was developed to provide transparency in our data production process. Although the NOFA requests a plain text readme file, we have expanded this notion to develop a more comprehensive deliverable.

Our goal is to illustrate a thoughtful process designed to meet the intent of the submission. Our hope is that we have developed a process that is reasonable, with respect to the data it deals with, as well as flexible enough to change with evolving NTIA requirements and lessons learned from the broadband mapping community.

Data Sources

Developing the Provider List

Provider lists for all states were developed prior to the first data submittal on March 31, 2010 from the following primary sources:

- State lists of regulated telecommunications, cable, and wireless service providers
- State and national industry organizations (i.e. cable associations, wireless service provider organizations, telecommunications associations)
- FCC Form 477 respondents
- Independent web searches
- Prior comparable mapping efforts
- Interviews with key state staff members and important community influencers

After March 31, we continued our research and added new providers to our list when discovered. In early July 2010, we once again initiated a telephone outreach campaign to contact all providers and to confirm the most appropriate contact person(s) within each company. Where necessary, we executed new NDAs with providers. This effort continued on a daily basis until we reached our final data submission deadline on August 15, 2010. As one would expect in a dynamic marketplace, this is an ongoing and important component of our work with respect to future data submissions.

As contact was made, we again verbally qualified each provider by asking a series of questions regarding the type of service and speeds offered. If the provider did not meet the minimum specifications for a broadband provider (as defined in the NOFA) we made a note of their status and removed them from the data submitted to NTIA. [1]

Provider Outreach

To meet the program's aggressive deadlines and participation goals, LinkAMERICA believes it is critical to maintain rapport with providers. To do this, we continued to reach out to providers with regular project communications, including a program newsletter and links to the various state mapping websites. We also launched early "private beta" versions of the interactive coverage maps to seek providers' comments and to include them in the development process. As described above, individual e-mails

[1] As with other Grantees, we have discovered a number of providers included in the first round whom no longer claim to meet NOFA requirements-particularly the 7-10 installation rule and/or resell services. Their data has been removed from the export to NTIA, but we do maintain the information for other purposes. As we explain in Appendix 1, our preference would be to retain the data but categorize them so as not to eliminate important providers. Also, since the first submission (data as of Jun 2009) new providers were discovered and added to the list.

and/or telephone calls were made to all providers explaining the status of the program and requesting their continued support in Round Two. We've also had the opportunity to support providers in their BTOP / BIP applications in certain cases. Through these collective outreach initiatives we continue to enjoy a healthy and appropriate relationship with broadband service providers.

NDA

To provide protection for all parties involved, LinkAMERICA continues to honor the terms of our NDA. If providers did not execute the NDA in Round One, they were giving an additional opportunity to do so in Round Two. New providers were of course also supplied with a copy of the NDA.

To facilitate the execution of NDA's, LinkAMERICA continues to use the DocuSign online document management solution. This system allows providers to review and digitally sign the NDA in a legally binding manner, and has been instrumental in achieving rapid approval and execution of NDAs with the majority of providers. In some cases, NDA's were individually negotiated to address specific provider concerns. In other cases, providers chose to submit data without executing an NDA.

Provider Survey

In Round One of the provider survey process, the LinkAMERICA team sought baseline information on network coverage and speed. That information was collected from most providers and was available as a starting point for Round Two – allowing us to collect changes from providers instead of asking them to submit entirely new datasets. However, because there were a small number of newly identified providers, we also chose to maintain the first version of the survey that allowed for complete submissions in multiple formats.

Survey Methods

Once again, we used a 100% secure digital survey process (via our provider portal websites) to collect and display information for providers. The Round Two survey process, however, was designed to accommodate both new and returning providers and the different types of information they would be submitting. The following is a summary of the process encountered by each group:

New Providers: New providers were routed directly to our standard survey where they were provided with templates for uploading data in tabular NTIA-compliant formats. As in Round One, if providers could not supply information in requested form, alternatives were offered. These alternatives included uploading service-area boundary maps, exchange area maps, CAD drawings, or customer address lists. From that information, the LinkAMERICA team developed a geographic representation of coverage and was able to build features for each provider.

Returning Providers: While many broadband providers submitted datasets in the Round One survey, most of those submissions did not contain 100% of the requested data. To help identify gaps, and to make the Round Two submission process as simple as possible, every Round One survey was reviewed for completeness, as well as accuracy and formatting compliance. Notes were made regarding gaps, and a customized instruction sheet was developed for each provider for Round Two. These sheets not only

explained what data was missing, but also provided directions on how to include that information in the Round Two process.

Check maps were also developed to show each provider how their service area would be displayed on the resulting interactive state map. Generating customized documents was an extremely time consuming verification process, but it allowed us to review provider submissions and close many of the gaps that might have otherwise persisted.

Follow Up

After the release of the Round Two survey in early July, LinkAMERICA launched an extensive effort to encourage responses. Every known provider was contacted at least twice by telephone or e-mail during the months of July and August. The initial data submission deadline was set for August 15, but we continued to accept "straggler" submissions well into September. We finally closed the door on new incoming data on September 10 in order to focus on formatting and normalization for submittal to NTIA. Any providers submitting data after September 10 will be included in a future update to the NTIA and statewide broadband maps.

No Response Policy

As mentioned above, every effort was made to contact each provider who appeared on our initial list. However, if no current information could be found on the company (i.e. no website, no contact person identified) and/or the provider failed to respond after repeated attempts, they were removed from the list. We believe the vast majority of those we were unable to reach were small wireless providers who have simply ceased to exist[2].

External Data Used

Beyond the data obtained from providers, we acquired the following commercial data products:

- American Roamer, Coverage Right Advanced Services
- MapInfo ExchangeInfo, Professional
- Media Prints Cable boundaries
- GeoResults Telecom Research Data

These commercial data sources were used to develop and cross-verify the provider list, estimate non-responsive provider coverage areas and verify the reasonableness of provider wireless coverage areas. We have tried to include third party data sources which touch on each of the three major technologies analyzed within the SBDD program. Each of these data sources ties back to a public domain data source which provides a cross verification mechanism for the commercial data product.

Although there are a large number of third party licensed data sources we remain conservative in our acquisition plans. From our limited analysis we are concerned about the ability to cross verify additional third party licensed sources against public domain data. Further we are unsure how we may be able to

[2] We will provide our response rates as part of our Quarterly Progress Report reporting cycle.

integrate another data provider's view of valid broadband providers within the definitions used by the NOFA (eg. are they using an FRN/DBA identity view or a marketing view, can the provider supply in a 7-10 day window, are they facilities based or the ISP only). This leads us back to a statement we made in a 'lessons learned' Webinar (April 2010) about exploring a consortia to lower the cost of data acquisition and allow multiple entities to peer review the quality and methodologies behind licensed data products.[3]

Beyond these commercial data sources, we used a number of public domain sources. These included:

- NECA Tariff 4 (Sep 2009)
- State produced exchange boundaries
- Carrier produced wirecenter boundaries
- US Census TIGER data[4]
- FCC 477 provider filers
- FCC Coals reports (321/325)
- FCC FRN lookup tool
- FCC/FAA Antenna Registration System
- FCC ULS Lookup Tool
- USAC Grant lookup tool
- USAC High-Cost FCC Filing Appendices
- HRSA data warehouse
- NCES data lookup
- State managed lists of schools (K-12),Post Secondary institutions and libraries
- List of museums, conventions, and visitors bureaus from www.onlineatlas.us

These public sources provided secondary coverage verification sources as well as assistance with isolating anchor institutions.

Finally, challenges exist when dealing with the inevitable conflicts between provider-submitted data and third party sources (public or commercial). There is no guarantee third party sources are more accurate or timely than the providers' own reports. Indeed, some third party sources are based upon different standards than the NOFA specified, perhaps making them LESS reliable than information collected via the SBDD. At the very minimum the provider data has a lineage and temporal status that we can identify. A concern we have with increasing use of third party data is we have no way to verify its quality or development methodology. In other words, we may hit a wall in which we can't figure out how the commercial source derived its coverage conclusion. To us this means that third party data sources are beneficial but they represent a supplementary view, not an authoritative one, of the NOFA defined Broadband market.

[3] We also suggested forming a technical standards committee and a consistent system for confidence reporting.
[4] Census data were derived from < http://www2.census.gov/cgi-bin/shapefiles2009/state-files?state=01>, Census 2000 files. Roads were derived from the county faces and edges file downloaded at the same location and tiled for a full state.

We have chosen to use provider data as the baseline. We will challenge provider reports when third party data shows major anomalies, or when a consistent volume of consumer feedback points to a potential error.

As the program evolves it is also our intention to provide tools that allow end users to evaluate the accuracy of the data in their own way. A confidence score or the presentation of multiple (and potentially competing) reports for the same location may be made available. This notion is discussed further in the "Validation" section below.

Confidentiality and the Use of Licensed Materials

As a mapping vendor, we are reliant upon the cooperation of broadband service providers. In large part, what underlies this cooperation is trust that we will not violate the proprietary and confidential nature of the data provided to us.

We are thankful for the confidentiality clarification that NTIA shared with us (included as Appendix Two). We intend to use this as a guiding document to help us communicate with providers what information NTIA considers to be confidential. Our suggestion is that NTIA publish this or something comparable, to ensure a consistent interpretation of the NOFA and how it guides NDAs.

As some providers are non-responsive to requests for information, or lack resources necessary to put data into NTIA compliant formats, we have fallen back to the use of commercial data sources in several places.

Some mobile wireless providers were unable to submit coverage information to us. In these circumstances we have generalized the American Roamer Advanced Services coverage. For incumbent telephone providers we have used commercial wirecenter boundary products to filter Census blocks which are clearly out of their exchange areas. Finally, licensed data from Georesults were used to derive estimates of broadband connectivity for hospitals within the Anchor Institution category. The value from Georesults was not used, but our estimate is modeled from their input data. We also use the name and address as provided by the State data provider, not Georesults.

Data Production Methods

As raw data were received from the provider community, attention turned to normalizing the disparate submission formats[5]. The team considered each submission with respect to the following criteria. These criteria are important because they perform the basis for our verification and quality assurance process. In other words, we have to appropriately scale our data verification efforts to match the scale or ambiguity of the following:

- Locational certainty

[5] In line with NTIA Best Practices we continue to request and receive a large number of data input formats. This ranges from tabular Block lists to hand drawn maps.

- Speed certainty
- Temporal certainty
- Provider and network ownership certainty

The team's goal was NOT to quantify a particular degree of precision with respect to any of these criteria. Rather, we are working to attribute the above "certainty attributes" to each submission, and will continue to implement quality assurance and verification mechanisms that are resource-appropriate for each.

Deriving Broadband Coverage Information

Broadband Coverage[6] was normalized into four formats:

1. Coverage in Census blocks (2000) of 2.00 or less square miles
2. Covered Street Segments (2000) in Census blocks greater than 2 square miles
3. Address Level Coverage (point data)
4. Wireless Service Areas (SHP file format)

With each submission, the team went through a series of steps to normalize and categorize the data. Since data arrived in many different formats, and at many levels of granularity, the following normalization procedures were used:

1. Determining the nature of service being provisioned (who is providing service and what technologies are in use)
2. Planning an attack strategy for the submission –understand the data, assign team members to various tasks
3. Geo-referencing the data; QA the georeferenced data
4. Geoprocessing the geo-referenced response
5. Segregating the submission into the correct NOFA-compliant submission formats

[6] Speed, Anchor institutions and Middle Mile facilities are discussed in later sections.

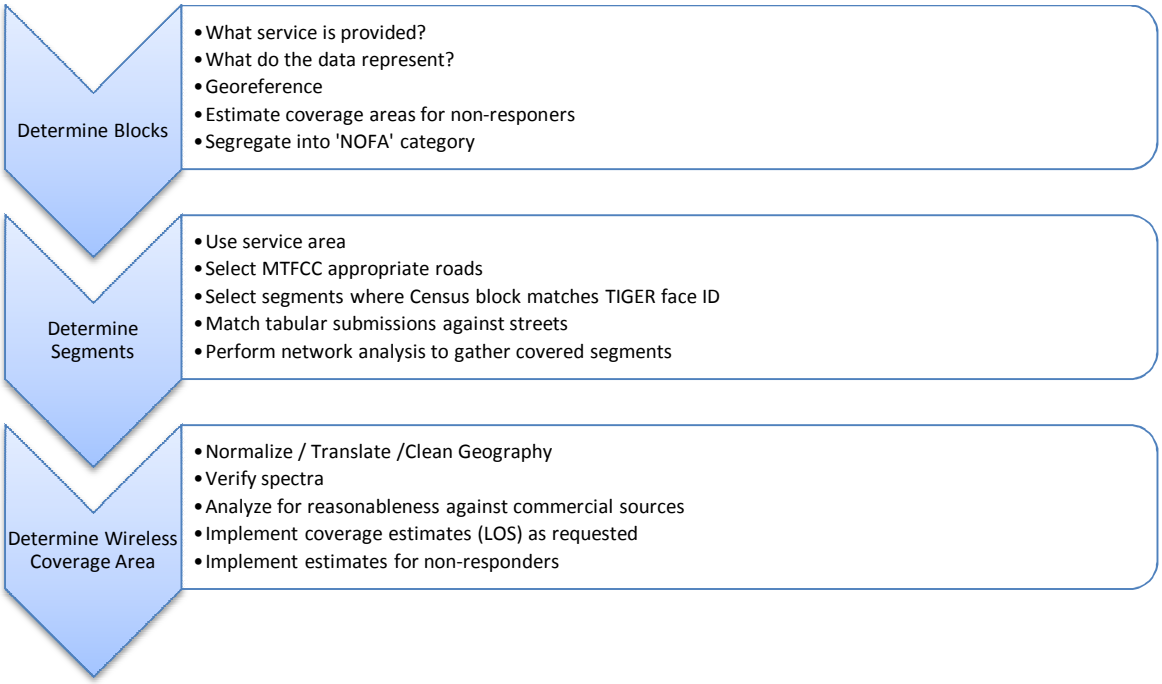


Figure 1-Broadband Coverage Process

Impact of Program Change

There were four important program changes during this submission that impacted how broadband coverage was developed and submitted to NTIA.

The first was the request to move Fixed Wireless coverage out of Census block/segment level to coverage areas more akin to mobile wireless providers. This was a logical change, but it impacted the data request that was sent to (and in many cases received from) this important class of provider. For the most part, the current submission reflects a movement of Fixed Wireless (Technology of Transmission 70 & 71) into coverage shapes rather than blocks. For a very small number of providers unable to provide coverage patterns and infrastructure to support that coverage, we are not submitting that coverage to NTIA. Consistent with our more conservative approach, which favors quality over quantity, our preference was to hold back the data rather than just provide unverifiable coverage patterns[7]. This hold back has created a small number of providers who are responsive, but responsive in a block level format which we cannot translate to wireless coverage shape files. For this handful of providers, we are holding the data but not submitting to NTIA at this time.

The second change was the choice to move from Census 2009 to Census 2000 geographies. Based upon our understanding of the first submission guidelines, CQA/LinkAMERICA developed an entire dataset built around Census 2009 geographies. Our online mapping systems were built based upon this baseline

[7] This could be argued as a bias against fixed wireless coverage whereas mobile wireless providers have rarely provided any infrastructure for their coverage patterns. At this time we have been using licensed transmitter points as a proximate check on coverage. Given the predominance of unlicensed transmitters in fixed wireless as well as transmission structures which tend not to fall into the FCC/FAA ARS, we feel comfortable with this decision. Hopefully, additional information can be supplied in future submissions.

information. Movement back to 2000 geographies forced the team to re-geoprocess much more information than anticipated in this submission. Rather than this dataset reflecting an incremental change upon the prior, this dataset actually represents an entirely new processing effort. Moreover, much of the initial verification and validation work was expressed in terms of certainty against 2009 block shapes. Because the shapes, sizes, and distances of the 2009 blocks are different than the 2000 blocks, much of this validation work has been (and is being) redone.[8] In terms of both time and resources consumed, this has been an extremely costly modification.

The third change was a request for a geographic object associated with covered street segments. This was a very reasonable request and we support it. As will be discussed below, a tremendous amount of processing has gone into trying to identify street segments based upon both non-geographical and tabular, or geographical references that have little certainty with respect to the derivation of the underlying data or its scale. Across four states, the number of providers who provided geographic segments is under six. The number who provided a geo-reference-able attribute (such as TLID), is probably in the same range.

The fourth change involved the introduction of a reseller category into the data. We have a number of carriers who we believe qualify as resellers (e.g., Covad, One Networks, Megapath), but as of this writing, there has been no program level definition for this category. Currently, these data are surveyed and retained, but held back from this current NTIA filing because they don't fit the broadband provider definition, and we are not sure what the reseller definition will be.

This class of 'held-out data' may impact response statistics.

Coverage Geoprocessing Methods

The next section discusses how data were geo-referenced and geoprocessed given a particular submission format.

In most cases we were not provided with any street segment level information for blocks greater than two square miles (large blocks). This necessitated subsidiary geoprocessing. As stated before our first goal was to derive block level coverage. Then, for blocks greater than 2.00 square miles, we moved to a segment gathering processing. The segment process will be described in the last section.[9]

[8] Much mechanized verification work has been done in terms of block distance relative to a service area boundary. When a block shape shifts from 2009 to 2000 (blocks at the 16th character are used in this analysis) the analysis result for a 16 character block object versus the dissolved 15th character block will be different. This makes it necessary to re-run the analysis against the appropriate block shape. Additional work with respect to block adjacency patterns also has to be thrown out for the same reason

[9] As has been discussed previously we note inconsistency in how providers are supplying information at the block and segment level. Beyond the temporal differences, we see that providers are computing area differently as well as including or excluding water areas. This provides an inconsistent measure across providers for the 2.00 sq mile cut off. Our preference would be to provide guidance to service providers within our states, but our concern is we will inconsistently message this with other grantees. We would like consistent guidance from FCC/NTIA on this prior to releasing notice.

Block Level Coverage Derivation Using Service Point Data

A number of providers submitted point level customer data. This number has increased from the first submission.

In some cases the submissions themselves were not internally consistent. For example, in the image below, unprojected points are shown, while the Census block polygon to which the points are supposed to "belong" is highlighted. In this case, either the block attribution is wrong, the points are not in the location to which they are attributed, or different block shapes were used than what is assumed.

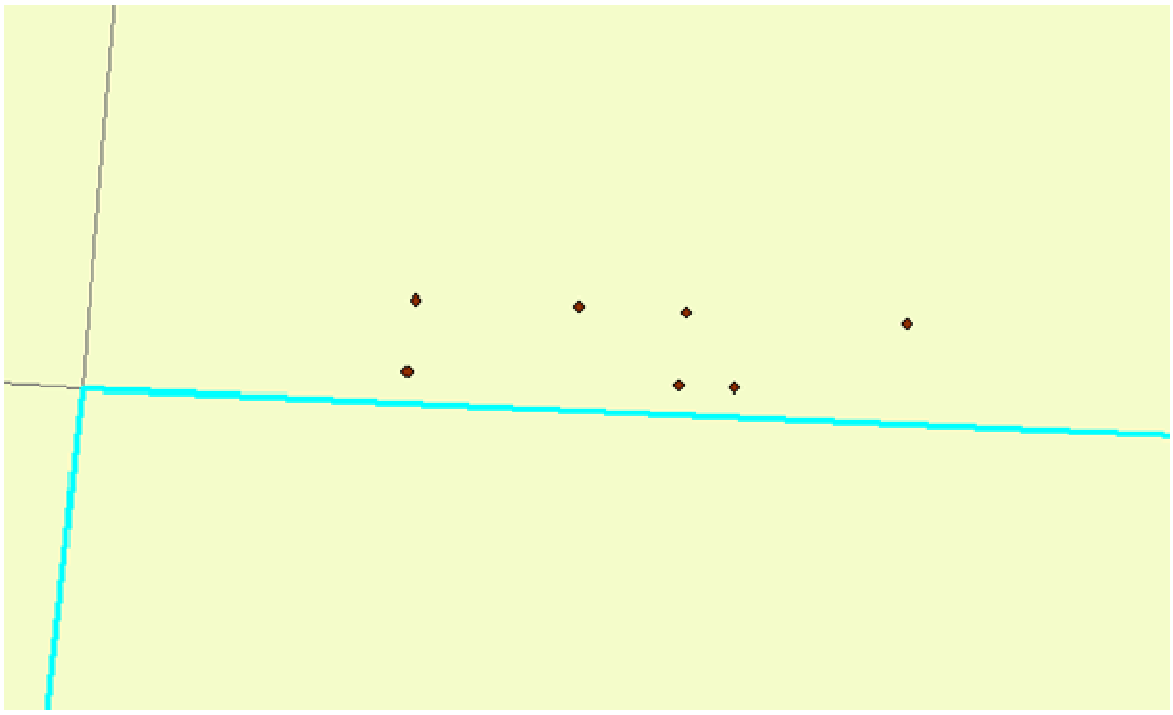


Figure 2-Internal inconsistency in submitted data

In other circumstances, we found that inconsistent geocoding standards may produce misleading results. The next image shows point level data, and the blocks are colored based upon the counts of points intersecting blocks. The challenge this presents is that if geocoding was performed on a different dataset than the block boundaries (the road traces are not coincident with block boundaries) and/or geocoding was done without an offset,it becomes problematic to assign coverage to a Census block based upon only the point locations.

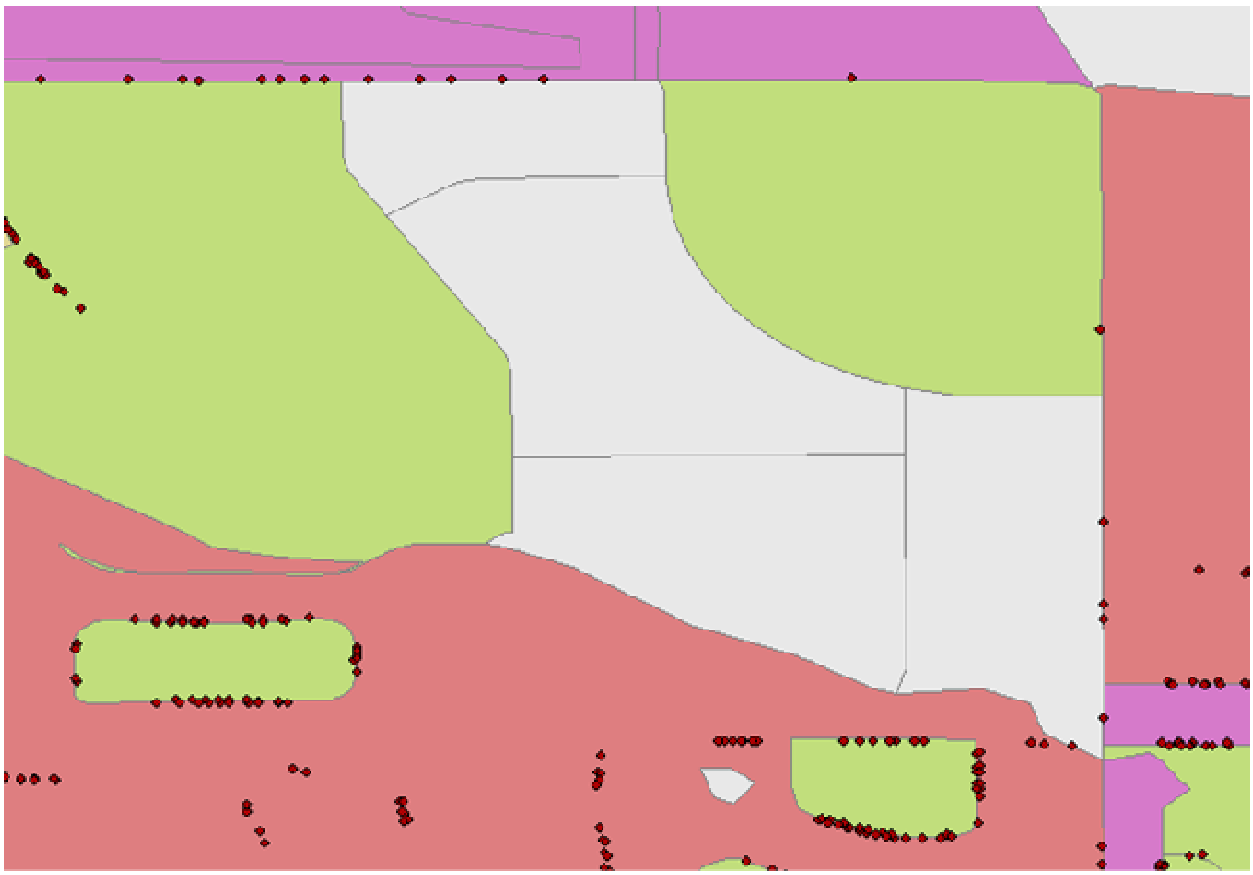


Figure 3-Block Coverage

For this reason, we elected to use a 200 foot buffer to select Census blocks that intersect our points.

Block Level Coverage Derivation Using Customer Facing Plant Level Point Data

In other circumstances some providers submitted point level plant data. From what we could gather, these points tended to be customer-dedicated terminals. Typically, these providers were high speed broadband producers—which may somewhat strain the definition of broadband as other providers supplying comparable services specifically disclaimed the ability to provide high capacity broadband services in the required 7-10 day interval. In these plant point data submissions we had similar concerns to the point level customer data, but two factors tended to make us use a more conservative intersection buffer. First, we tended to have far fewer points to work from, so our concern was grabbing too many covered blocks as the blocks tended to be much smaller in these urban areas. Second, these plant points tended to be dedicated to distinct customers, but it was difficult to know which element of the customer's campus to attach coverage to.

In the case of the image below given a small shift to the left, it would be easily possible to gather 1 to 3 Census blocks from this point. Although orthoimagery is helpful in a circumstance such as this, it is still indeterminate – specifically in areas where the coverage is attributed.

Thus, in the circumstance of plant level point data we used a 100 ft intersection buffer.

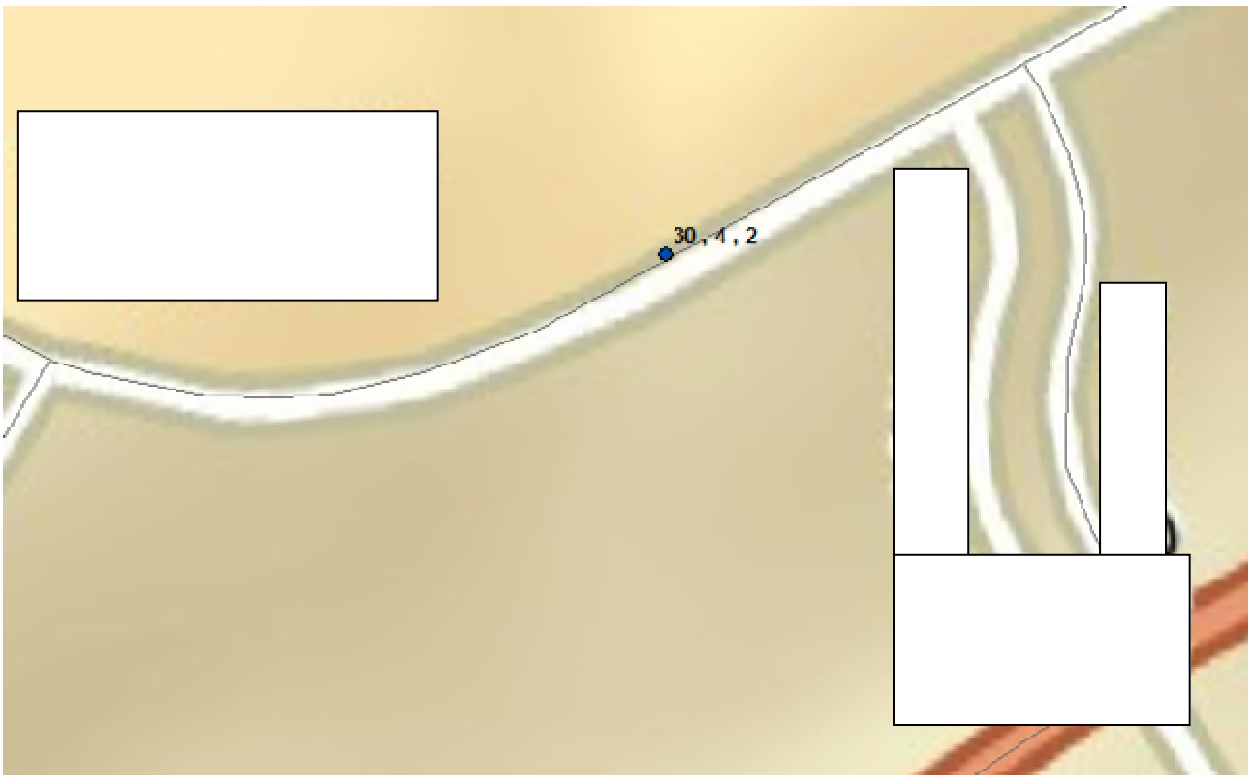


Figure 4-Plant Point level data

Coverage Derivation Using Linear Facilities Data

A number of providers submitted facilities data. We handled this data in different ways depending upon what we believed the facility data strand represented.

Most telecommunications networks are divided into two components. Feeder supplies higher capacity nodes (eg DSLAMs, Fiber Nodes). Distribution usually supplies customer premises (NIDs, Pedestals, Taps, ONTs). Where we could discern what strand we were provided, we used different methods.

The next image demonstrates a geo-referenced CAD image as given to us by a broadband service provider. Note the light and dark green shading. We would infer that the lighter segments represent distribution and the dark green represents the feeder network.

In the case of a combined strand map, we used a relatively tight buffer of 200 feet to gather covered Census blocks. Our intersection tolerance is based upon an assumption that our data likely represent a situation comparable to customer point level submission in that we have most of the network footprint captured.

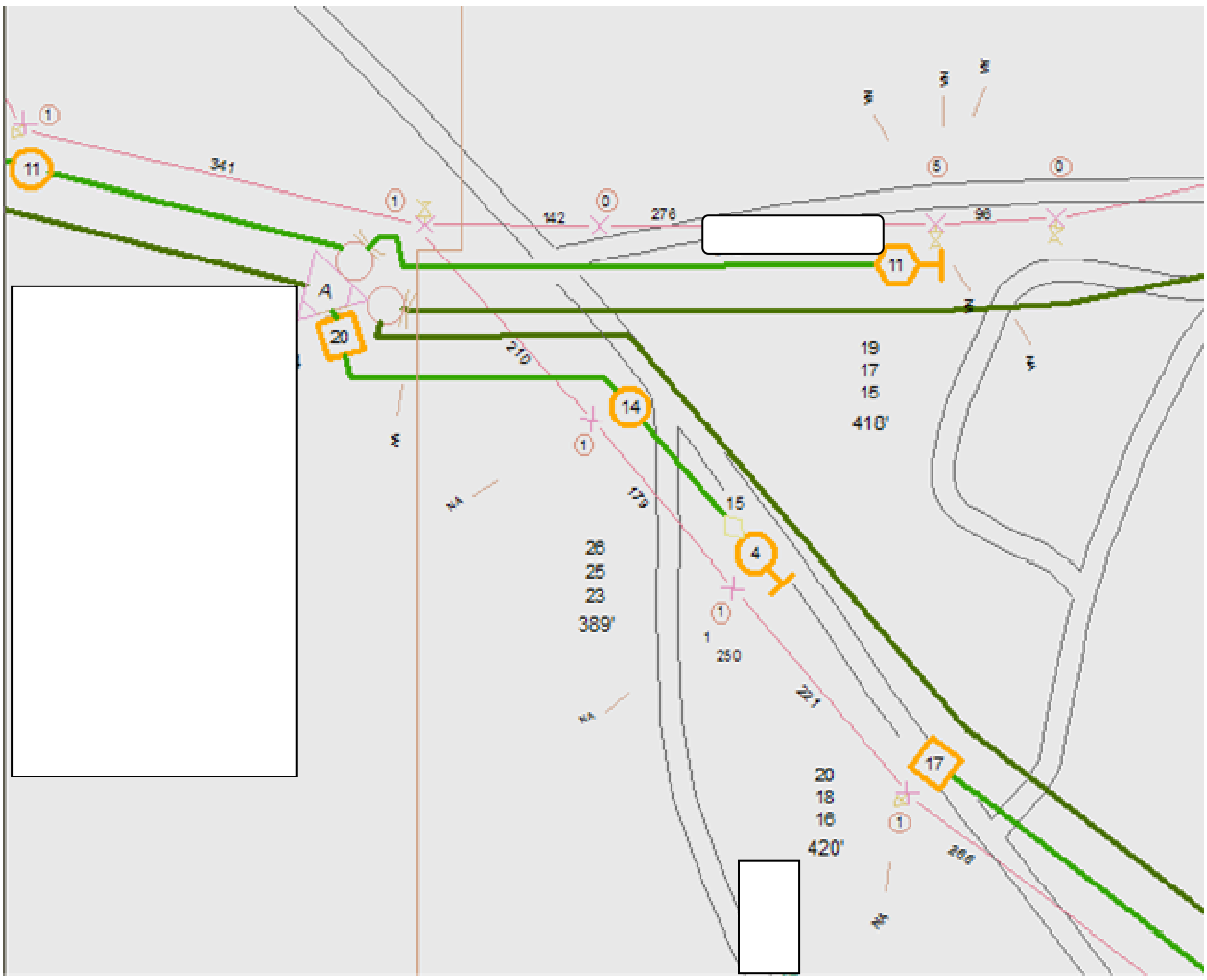


Figure 5-Georeferenced CAD information supplied by Broadband provider

In other circumstances we were provided engineering information which we inferred to be feeder only. This inference was typically based upon the presence of fiber optic equipment only. In these cases we used a more generous 2000 meter Census block intersection. The 2000 meter criteria was based upon an informal survey of population in proximity to the geo-referenced strand data, but it could be varied based upon a more complete survey.

Coverage Derivation Using Covered Street Segment Data

In some cases we were provided with covered street segment data. Covered segments tended to come from two sources.

In some circumstances providers gave us CAD data which was not drawn in a projected manner. This is relatively common for older engineering data derived from hand drawn records. This meant that our team had geo-registered the image into an approximate position. In this case, the boundary streets

were selected, and an enclosing polygon was derived. The intersection of this polygon and the blocks within became the geoprocessing method to derive blocks.

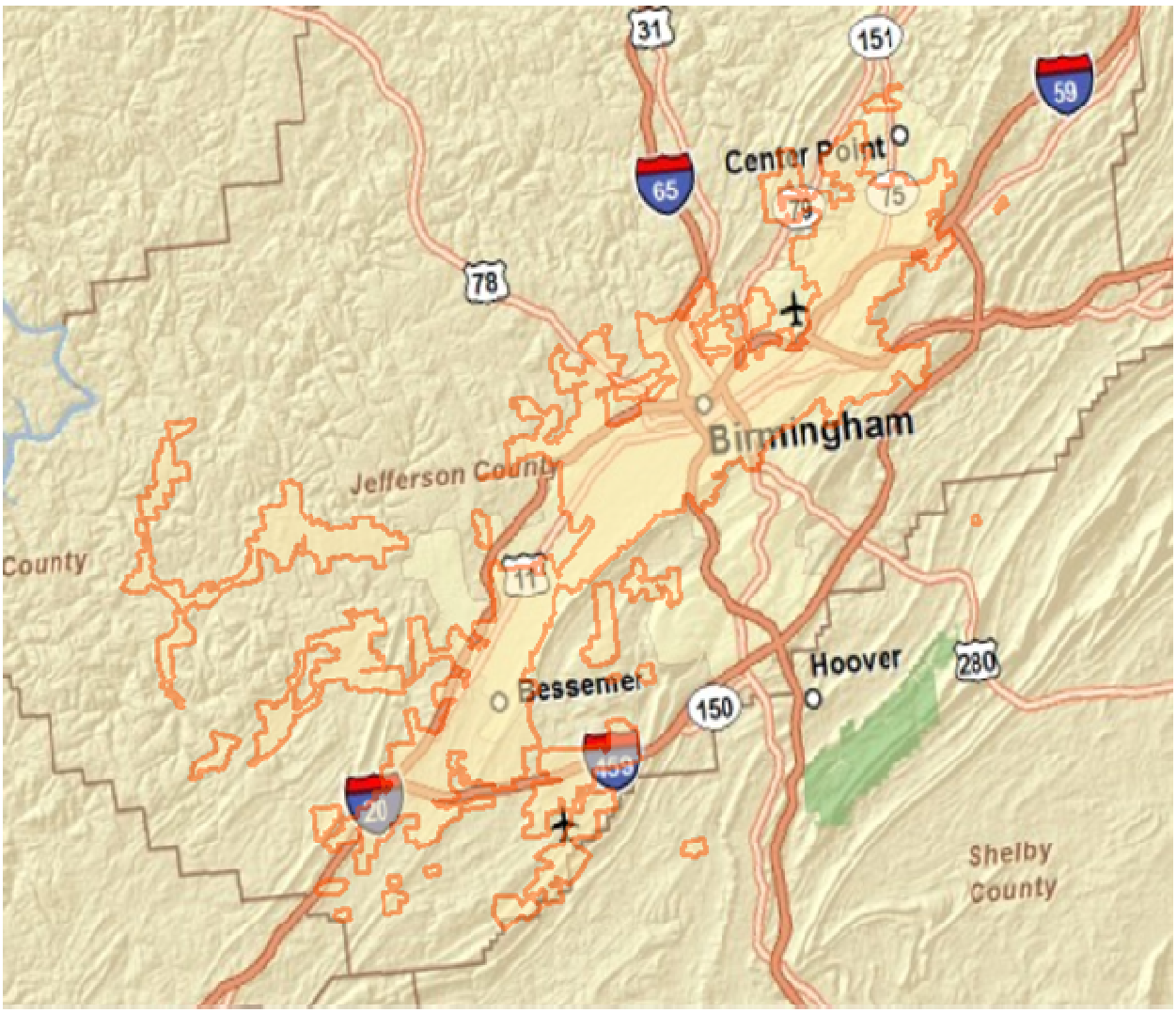


Figure 6-Coverage derived from street segments

In a second circumstance, street segment data was developed during coverage estimation. Handling the estimated data is discussed below.

Coverage Derivation Using Serving Area Point Submission Data

In other cases we worked with a provider to derive service areas based upon point plant data. In these cases we were given a primary serving node and an appropriate road length service boundary. There is an important distinction from the plant data discussed above. In this specific case the data submitted was a node that served many locations-such as a Central Office or DSLAM. This is contrasted with the earlier example in which the point represents a node serving only a few customers.

When trying to derive coverage from Central Office or DSLAM nodes, the team used ESRI Network Analyst to derive covered road segments honoring these road engineering parameters.

The figure below shows street level coverage derived from Central Office and remote DSLAM point data.

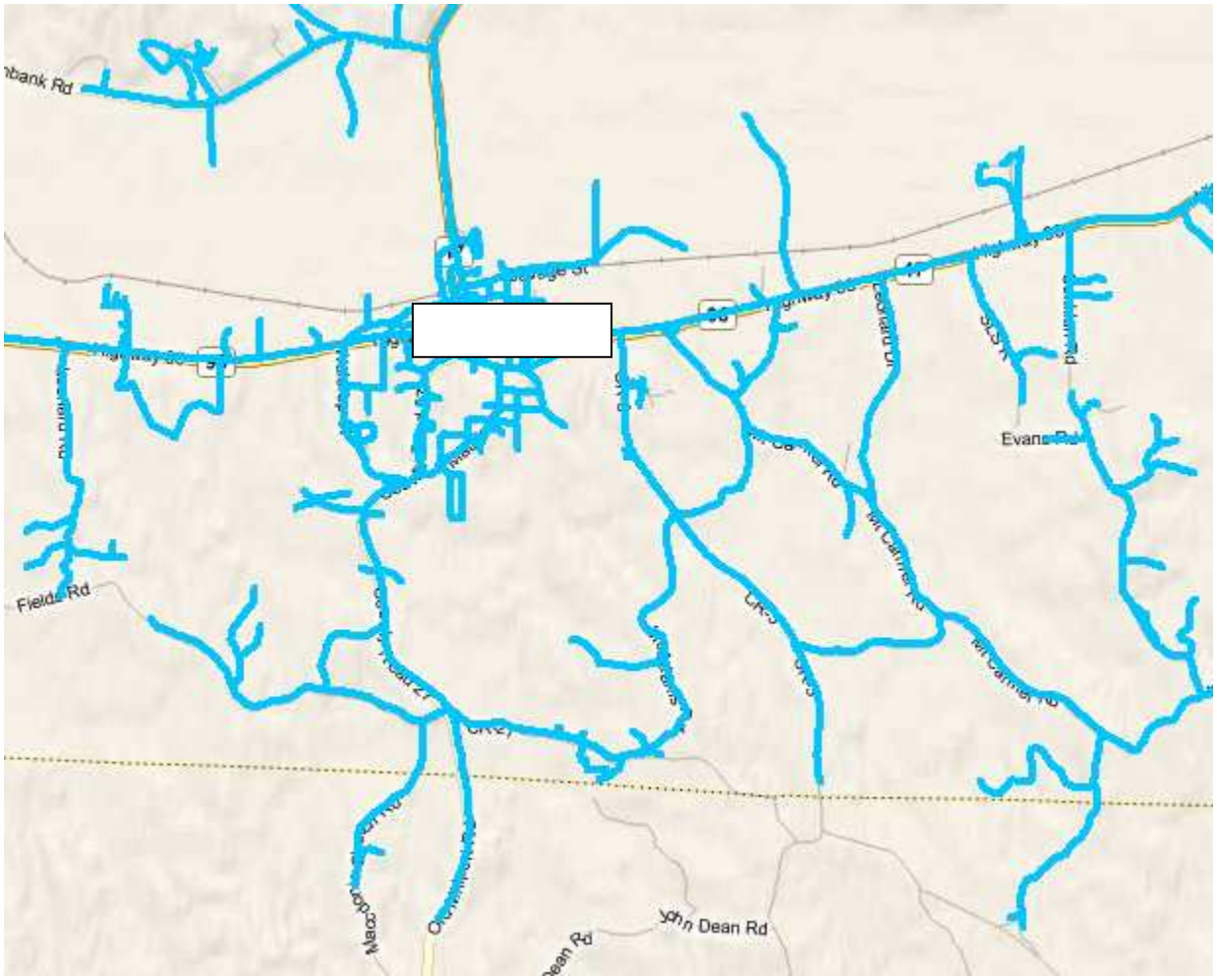
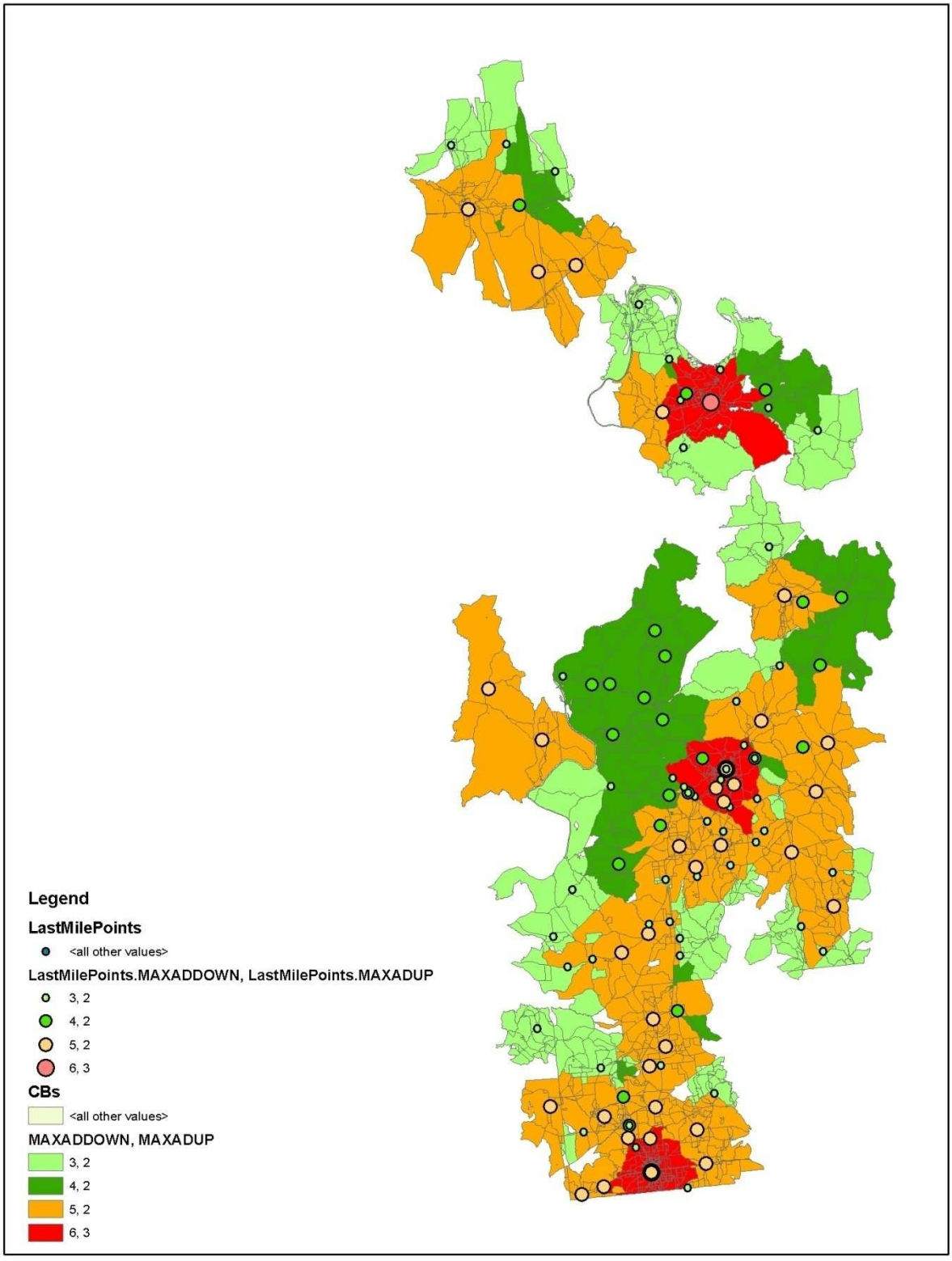


Figure 7-Coverage derived through road paths

In response to Provider feedback we revised this process to include a larger variety of TIGER road types. In submission 1, unimproved roads were not used. In this submission-particularly to improve estimates in areas bordering parks and public lands-a wider class of TIGER roads was used.[10]

The segment level coverage is easily extendable to derivations of Census block level speed. The figure below shows the attributions of block level speed based upon the Maximum Advertised Speed available from a DSLAM. Although the methodology isn't perfect, it does provide insight into the value of granular infrastructure data.

[10]Only TIGER features of MTFCC type S1100 and S1200 were excluded from use.


Legend
LastMilePoints
<all other values>
LastMilePoints.MAXADDOWN, LastMilePoints.MAXADUP
3, 2
4, 2
5, 2
6, 3
CBs
<all other values>
MAXADDOWN, MAXADUP
3, 2
4, 2
5, 2
6, 3

Coverage Derivation Using Polygon/Polyline Serving Areas

Broadband service providers sometimes submitted coverage in terms of served areas. This was either in direct geospatial formats, CAD files, or paper maps. The image below reflects a carrier's service area. Within that service area there are variations in technology of transmission and served speeds. When polygons with speed data and technology of transmission were available, we used a spatial intersection to gather covered Census blocks. In many cases, using covered Census blocks resulted in a loss of the speed variation (sometimes the speed variation was at a level below a block and doesn't get picked up within an intersection query).

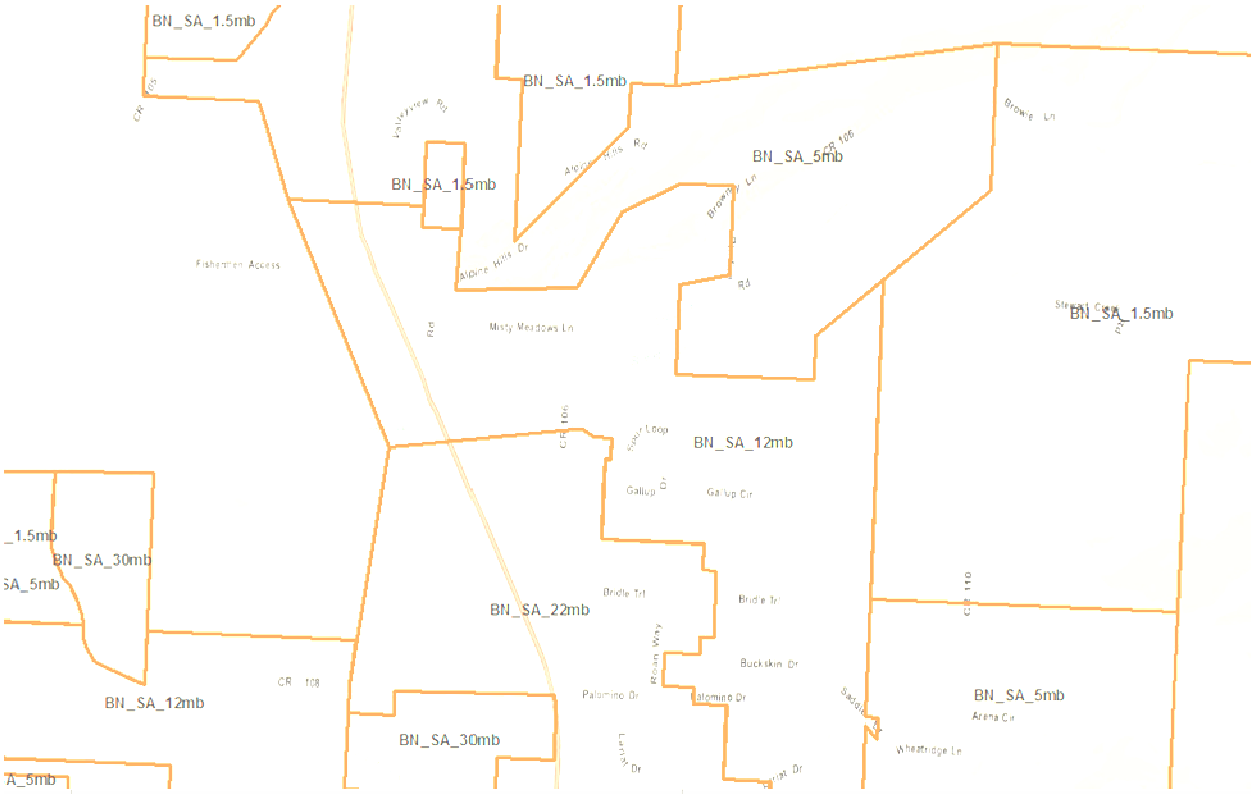


Figure 8-Coverage derived through serving area polygons

Although we cannot directly solve the loss of speed granularity due to Block shapes, for this second submission, we have instituted a business rule wherein we always select blocks from the highest speed areas first, and then allow the lower speeds to select from the remaining blocks. This is an arbitrary rule, but our feeling was that it should be a consistent selection, rather than an unordered selection.

Street Segment Derivation, Large Blocks

For those calculated blocks greater than 2.00 square miles (large blocks), we provided coverage in terms of covered street segments and corresponding geography.

With respect to segments we had four sources of data:

1. Covered large blocks
2. Tabular street segments and address ranges for large blocks

3. Geographic segments either with street attributes or without.
4. Service area boundaries

A number of providers only provided a list of covered large blocks without corresponding segment information beneath the block. This provided the dichotomy of either selecting all segments in the block, or none. Because we had little information from which to make the selection, we elected to be conservative and did NOT pass any covered segments to NTIA from this submission format. Some broadband providers submitted covered street names and street ranges. In these cases we performed a manual analysis trying to link to specific segment names and address ranges within covered blocks. Sometimes this was a simple process because a provider used a TIGER derived street database. In other cases we could not determine the source of the provider's street data . Street and Address matching tended to yield a relatively good result (typically between 50% and 100% of possible segments in the block), but was very time consuming. Where yield rates were low, our result was a shredded segment coverage pattern, like the image shown below.

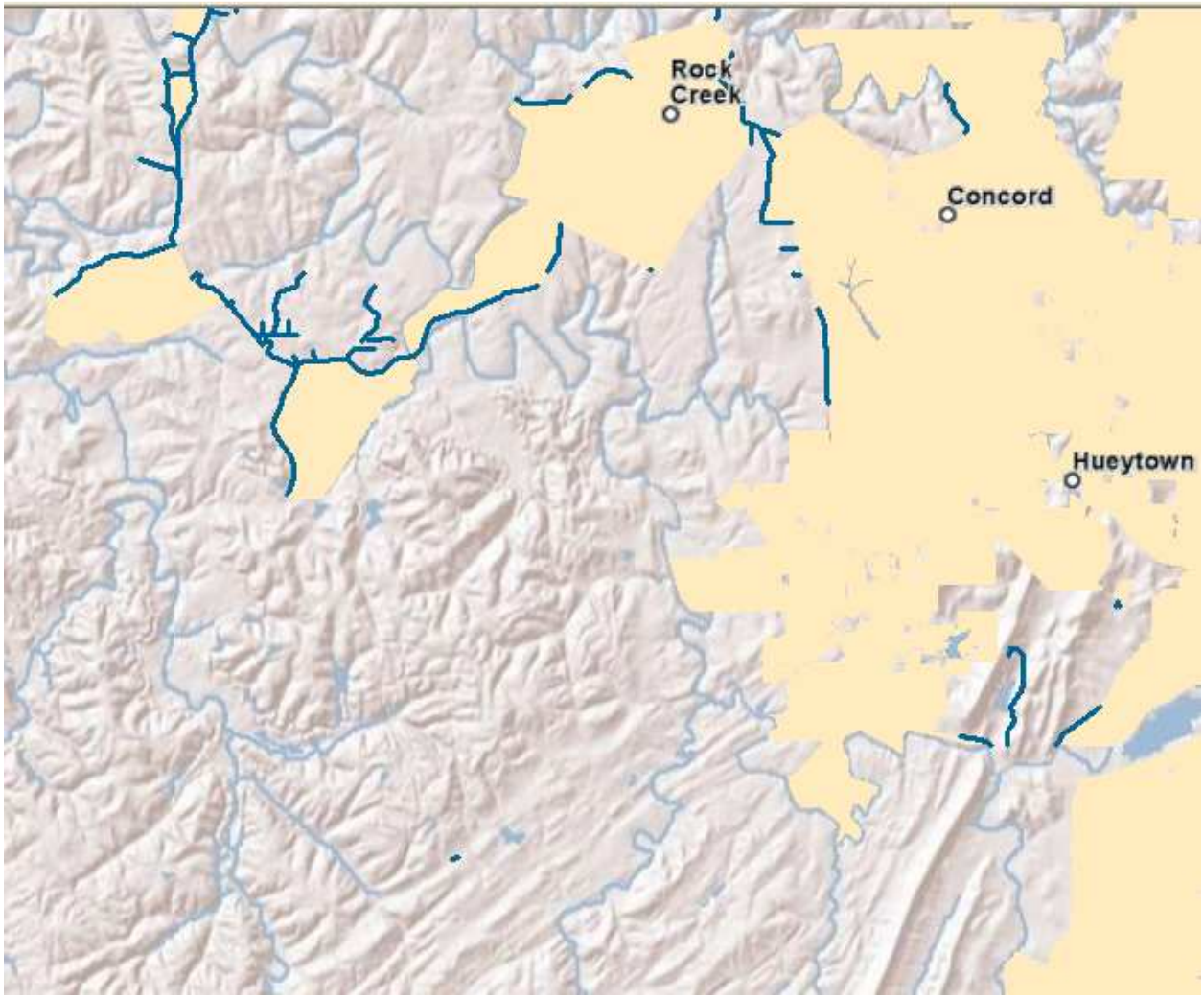


Figure 9-Blue road segments adjacent to peach covered small blocks

A number of providers submitted geographic objects. In this case, our manual process was directed toward a conflation of data sources. The goal was to take provider submitted segments and put these segments in terms of our TIGER 2009 basemap. Although there is a trade-off in the accuracy using non-provider submitted segments, we felt it was more important to have a road set that would edgematch our block features. This is important for the appearance of the online maps, as well as potential verification work where we are attempting to judge a feature based upon its attachment to a covered small Census block. The figure below shows street segment input data.

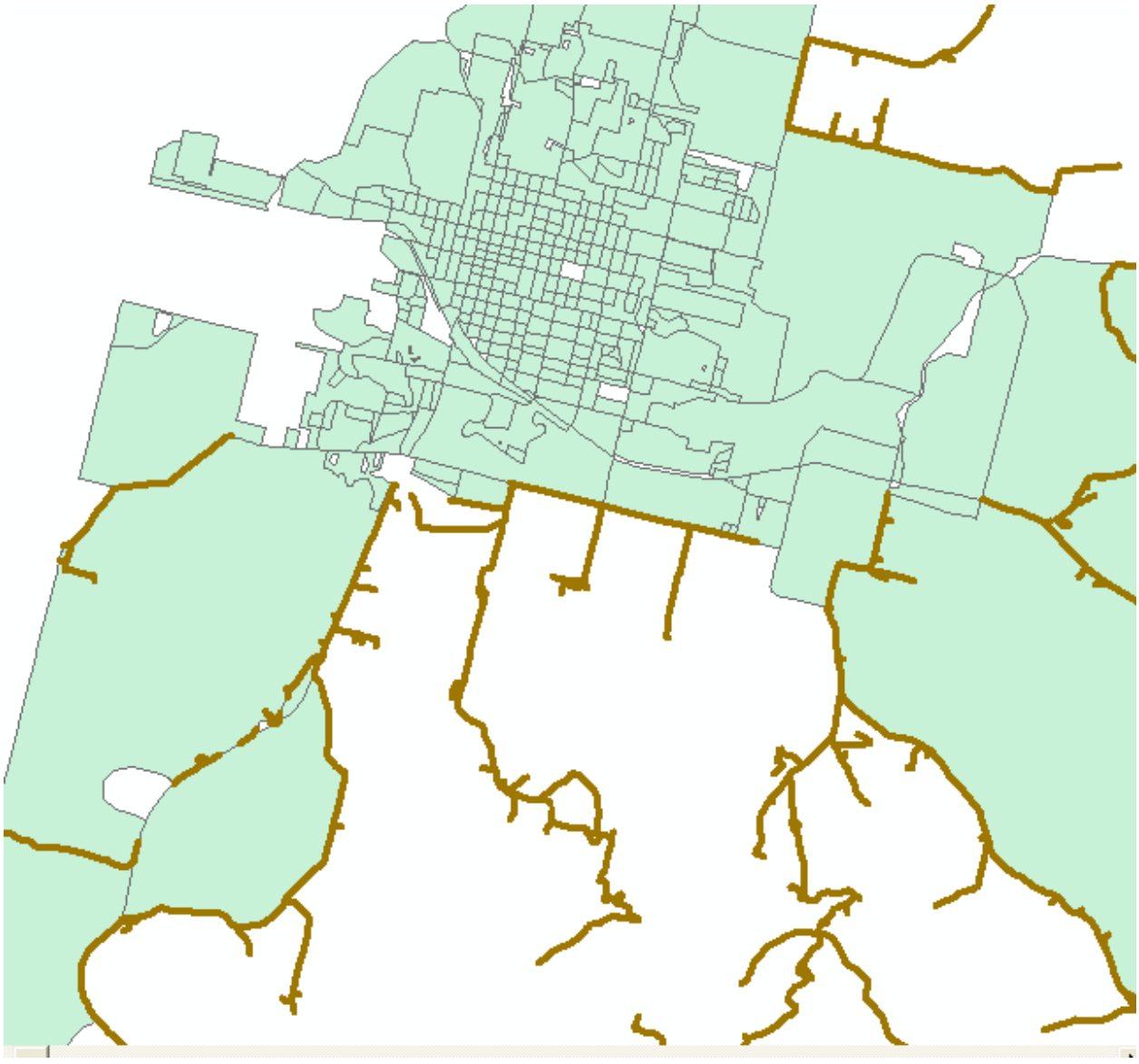


Figure 10-Provider Submitted Street Segment Objects

The figure following demonstrates the same area after the conflation process. Blue segments are the conflated TIGER roads which will be passed to NTIA.

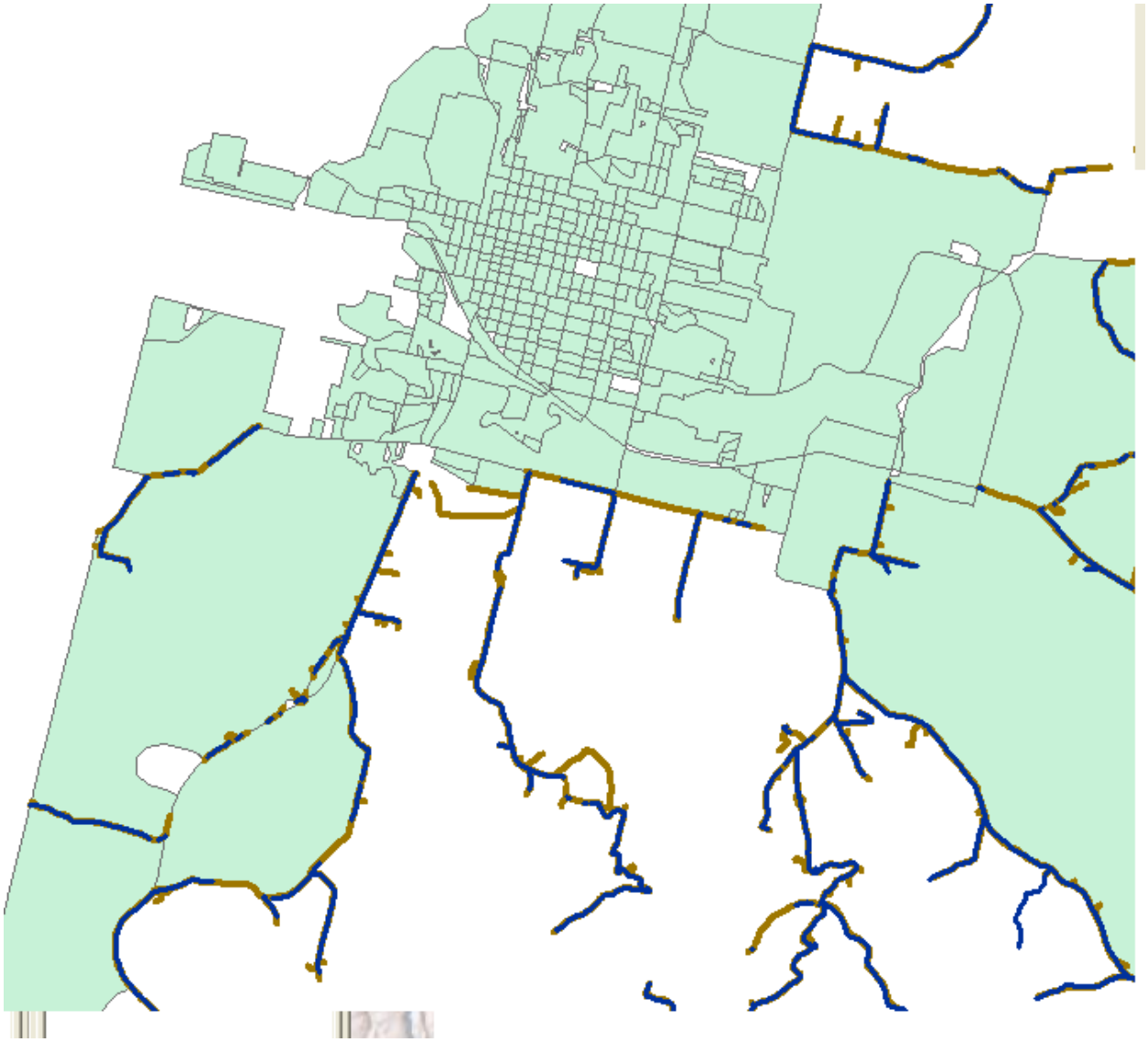


Figure 11-Provider submitted segments in gold, selected TIGER 2009 in blue—Conflation result

The final segment process was used when we were supplied with a broadband covered area polygon. In this case we knew the bounds of coverage so the effort involved finding the segments within covered areas and eliminating those segments inside of blocks less than or equal to 2.00 square miles.

Because there was more control over the format of the inputs (we knew we had a boundary and were working with TIGER segments), this was an automated process that followed this general format:

1. Select large covered blocks by provider ID (from updated Large Block table)
2. Select TIGER 2009 road segments (MTFCC like 'S%') that face (CB = CBLeft2000 or CB = CBRight2000) covered large blocks for provider
4. Select segments as distinct records, max speed with corresponding technology,join in feature names, Export selected records to temporary DBMS table

5. Join TIGERroads feature class to temporary table on TLID
6. Select covered segments (Python script)
7. Select service area polygons for provider
8. Clip selected facing segments with selected service area
9. Export clipped segments to staging feature class, keyed by ProviderID

In this figure, orange represents covered small blocks; black lines are covered segments in large Census blocks (light blue). The service area boundary is shown in grey. Based upon feedback from providers, we have elected to Clip segments at the end of a coverage boundary.[11]

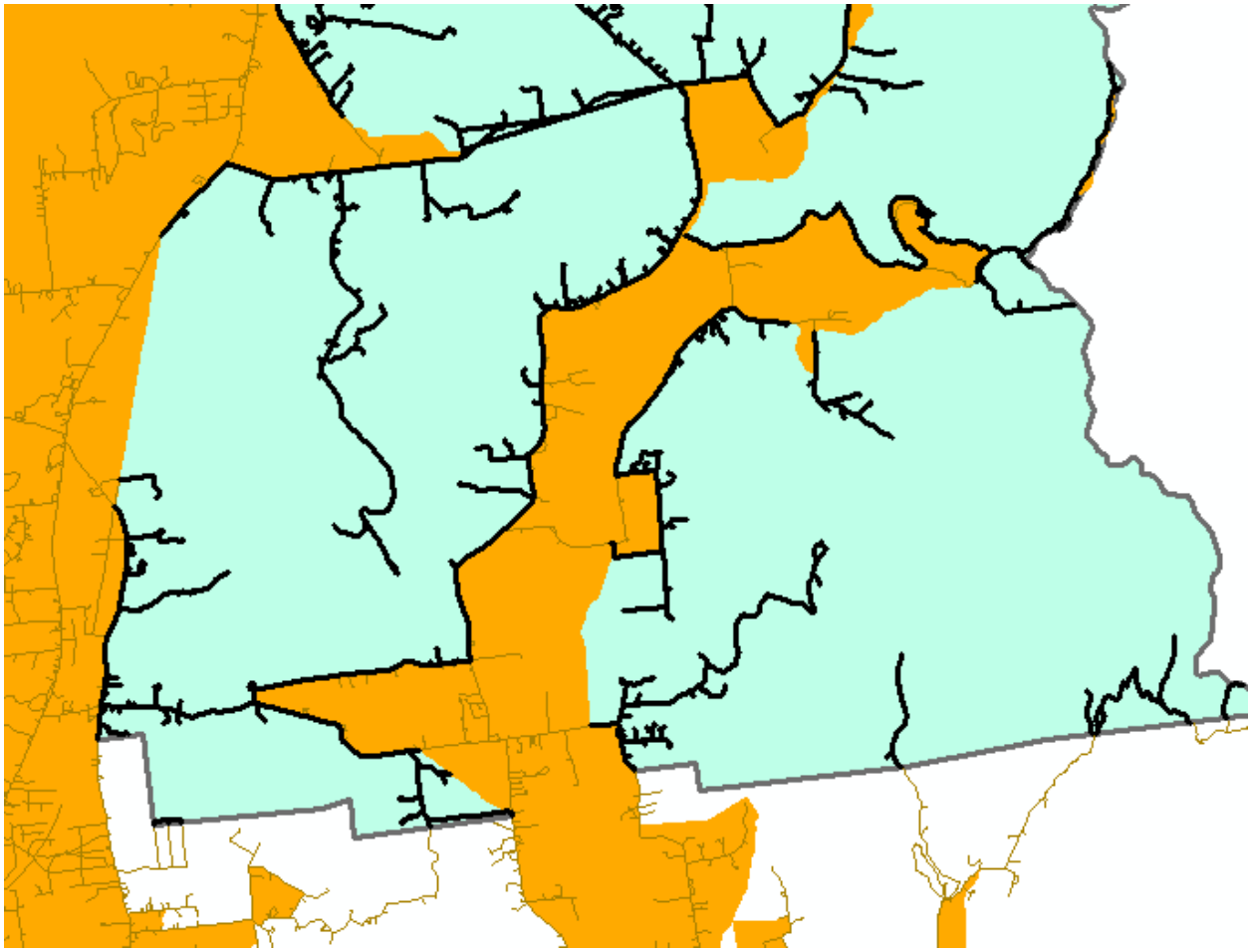


Figure 12-

Wireless Coverage Process

In general most providers of mobile broadband submitted coverage information in a NOFA compliant format. Other than attributions for spectrum and speed, little was done to this coverage.[12]

[11] An outcome not discussed here is how to handle address ranges on segments. As NTIA is asking for a Min and Max on the segment, deriving theses values for clipped segments is very problematic. Also the prevalence of alphabetic characters in addresses makes the min/max selections very arbitrary. We are grateful that addresses are nullable data elements.

In large part, fixed wireless providers either supplied coverage information or infrastructure from which coverage estimates could be derived.

Many fixed wireless providers allowed us to use their tower locations, antenna heights and direction/spread of coverage to derive a line of sight coverage estimate. In our experience, this is a conservative and reasonable derivation of coverage.

Some wireless providers submitted RF studies. When this was done, there was a request that the signal strength be removed from coverage data. The request was honored.

We do note two interesting trends in the wireless data. First, there appears to be some variation on how the NOFA coverage definition is met. In other words, there seems to be a disparity on the necessary strength (e.g. -80 dB, -98 db, -120 dB, etc) to provide the appropriate quality of service for data services. Second, it was very difficult getting providers to identify spectra used for broadband data services. We are unsure if this is a competitive concern, or if the same coverage pattern is yielded for multiple frequencies. Typically, the spectra returned were those that a provider was licensed for. At this point, we have no reliable way to locally determine what set of frequencies are used to provide broadband data services in a local area.

Service Address Point Process

A handful of providers have requested that customer level, service address point data be submitted to NTIA. In these circumstances we have done minimal processing to preserve the provider's intent with this deliverable and not bias downstream NTIA use.

Our verification included checks against commercial or Public Utility/Public Service Commission exchange boundary maps. Points not contained within 1 mile of a boundary are not submitted to NTIA.

We retain from the provider, the provided latitude and longitude, as well as Census block. We test for comparable geocoding success rates but do not overwrite provider information. From this analysis we note the amount (usually more than 10%) of addresses that seem to locate with less than street segment certainty. Deriving a thematic representation of the points on speed, also illustrates some of the locational certainty issues in this point level data.

Coverage Estimation Process

Although the derivation of Broadband coverage into Census blocks, street segments, or wireless coverage files is, in itself, a bit of an estimation process, there was an explicit estimation process required in cases where a broadband provider either refused to participate in our survey, or provided such a threadbare submission that no carrier-based coverage information could be gleaned.

We typically resorted to three possible estimation paths.

[12] Some polygon data did exceed the Oracle node count threshold. In these cases, data was rasterized to 100m cells and the converted back to polygons. The polygons were dissolved to multi-part geometry. This addressed the node count concern.

For Cable (HFC) providers who did not provide any coverage information, we fell back to Media Prints data. Rather than using the entire Census block group gathered by Media Prints, we used only those Census Designated Places carrying the same or similar names to the Media Prints p_com field. Our reasoning was that Cable systems tend to be franchised on a municipal or at least administrative basis so the coverage will likely follow a governmental boundary. As a general rule cable infrastructure is not available in the public domain[13] and what could be found was poor in quality and difficult to ascertain for validity.

For DSL providers who did not provide any coverage information, we estimated road-based coverage from their Central Offices[14]. We only used Central Offices that showed evidence of DSL or fiber-based services in the NECA 4 tariff. Road-based engineering areas were derived via ESRI Network Analyst to 18kft. These segments/boundaries were clipped to commercial wirecenter boundary edges.

For mobile broadband providers who were non-responsive to our requests, we fell back to American Roamer coverage patterns. We generalized the American Roamer coverage to ½ km in order to protect the licensed information.

For fixed wireless providers who provided no coverage information, we relied on their public websites to glean coverage maps. When these maps were available, we georeferenced them and tried to use the outer polygon boundary to represent their serving area. In other cases when only a tower could be provided, we used a view shed analysis and estimated coverage at 10mi per tower[15]. As much wireless propagation is driven far below the Census block and much engineering information isn't known (frequency in use, polarization of the signal, coverage pattern of antenna(s), local terrain/land cover) this was the most complicated group to estimate. At this point, we have done this with only a handful of providers where we could test the reasonableness of the estimate. The remaining coverage areas are generally inconsequential to an overall view of state broadband coverage, even though they can of course be more significant within the context of a more localized planning process. We continue to work on this class of providers.

Speed

Speed attributes are reported both at the block (typical) and higher levels (maximum advertised and subscriber weighted). We note that in many cases, providers did not supply typical or subscriber-weighted speeds. In some cases it appears-although we cannot verify-that their maximum advertised speeds were used in typical speed columns.

[13] The team tried to use data from the FCC Coals system and 321/325 fillings but this seemed to be a bit non-uniform in quality.

[14] Central Office location was derived from MapInfo ExchangeInfo Professional, September 2009. Wirecenter boundaries also came from this commercial product.

[15] In some cases we had an approximate radius of coverage but no height. In this case we used a 50' height estimate and then clipped the coverage to the provided coverage range. We also clipped wireless coverage to honor state boundaries but did not look for providers serving coverage with out of study state facilities.

We do have limited testing data on reported speeds, but we have been careful to not use our typical reported values with carrier-provided information. If we do not have a speed value from a provider, we report an empty string.

Many service providers claim they do not have data on typical speeds available, but estimate a 20% overhead factor between the advertised speed and what may be experienced by an end user.

During the second survey we did request advertised speed at the block level. We appear to be getting speeds that do not vary over a large geographic area – leading us to believe providers may still be submitting the maximum speed advertised in local media for the entire market. For the most part, we have been unsuccessful messaging the recent description that advertised speed should not correspond to a market area, but instead, the maximum speed which can be provided to a household—what some may describe as a qualified speed.[16]

Community Anchor Institutions

In the first submission, the Community Anchor Institution (CAI) process was referred to in terms of a learning curve. This continues to be an appropriate metaphor. The mapping team continues to focus on data that will support and help inform policy makers and the SBDD planning process.

In the first submission, the team gathered information on what data was available and what resources will be required to engage these categories of important institutions. In this second submission we continue to obtain additional connectivity information, but we also look at the CAIs as a primary leverage point with and for the planning process.

Our work with CAIs is guided by three principles.

First, CAIs are important stakeholders within the planning process. Our goal is to engage participants in regional planning that have strong ties into the eight CAI categories identified by NTIA. This has a direct benefit of engaging an established stakeholder community. It also allows Broadband planning to tie into existing organizational and planning networks. In each of our states, key relationships with education, public safety, libraries, and economic development sectors are being identified and developed.

Second, we believe that CAIs will likely be one of the primary beneficiaries of targeted broadband funding. Our belief stems from the sense that many of the benefits of broadband will extend from these community 'anchor points'. In other words, it isn't solely the existence of broadband at a library that provides a benefit. It is people using applications that work only on a broadband network to upgrade their skills (e.g., online training) and gain access to online content (e.g., job postings, goods and services), etc. The targeted use of a specific application--that can only take place with broadband

[16] As an example of a response to our request for Block level advertised speeds, we received the following comment from one anonymous provider, "This is and of itself does not require anything new of us – just states the NTIA supports efforts focused on getting that information on the CB level." It would be helpful to have broader messaging so that providers understand this new direction.

networks-- is what produces the priority benefit. Put another way, there seems to be a realization that things are less about pure connectivity (for the sake of connectivity) than about connectivity in terms of an application (for the sake of the benefit obtained through the application).

Third, we continue to use a rational and targeted approach to derive information. This means we will utilize our planning teams for as much ground work as we possibly can. This also means that a goal of our CAI process is not an exhaustive Census of anything that could be a CAI; rather it is the discovery, inventory, and integration of broadband planning activities into those CAIs which stand to produce the greatest synergies with the SBDD planning process. This implies two significant points. First, the team's goal is to document community anchor institution connectivity within a broader context of regional and statewide planning objectives. Second, if a particular category of CAI has an independent broadband planning effort underway, we will encourage that organization to take the lead, and we will provide relevant expertise and support as warranted. For example, in one of our states, the public safety community is already engaging in a mobile broadband survey effort. We have aligned our CAI data collection process with that effort and are sharing information and expertise (e.g., hosting a survey) to support their mission.

Further, the team continues to rely on the notion of Internet Intensity Zones. As the broadband coverage information is developed in this study, those Anchor points that fall into an existing area of SBDD broadband coverage will not be surveyed at a detailed level. Rather, the adjacent coverage area will be the first estimate of broadband coverage for the facility. The use of an estimate allows the site to come into the analysis and learn a bit about the accessibility of that facility, but it also frees resources to examine those anchor points that are more dispersed and likely under/un-served. The team will conduct targeted surveys to discover connectivity and, more importantly, applications in use at prioritized CAIs.[17]

We close this section with a figure that we hope reinforces our CAI process.

[17] We track internally those features with Broadband connectivity defined via an estimate but within the current transfer data model we lack a mechanism to propagate that information to NTIA. Appendix One expands upon our thoughts regarding a series of audit fields in the transfer database which would be helpful to inform downstream users regarding the source of data or use of estimates.

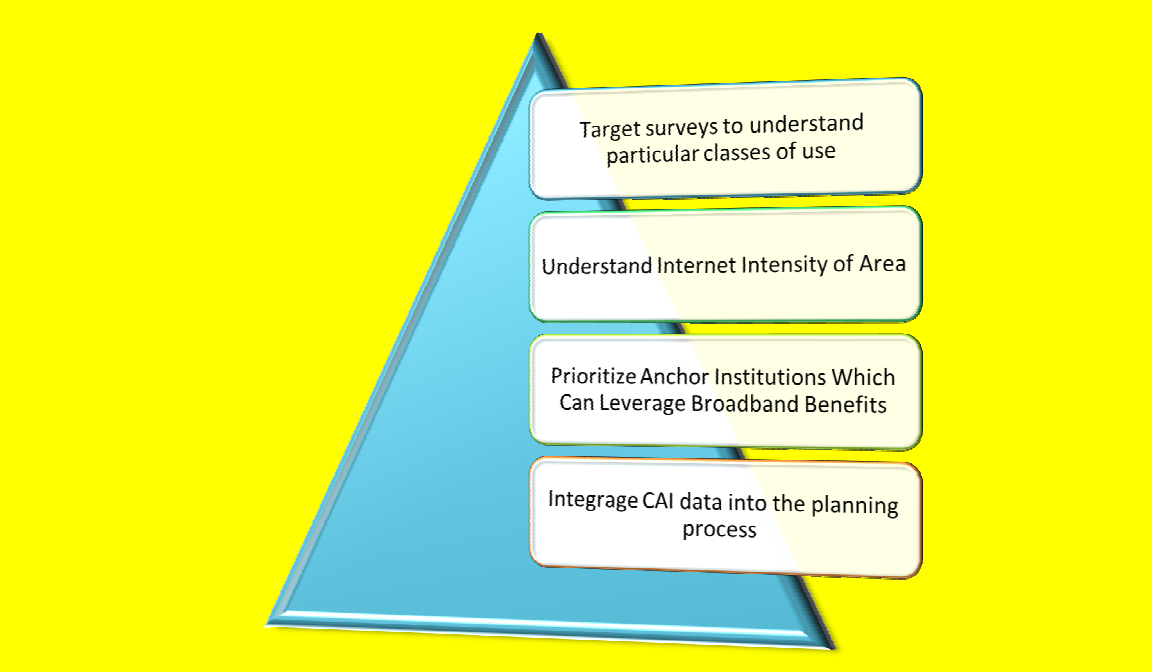


Figure 13-Anchor Institution Process

Recall from our first submission analysis, in most cases, CAI points are clustered and on average less than 1 ¾ miles away from one another. Relying on The First Law of Geography[18], this likely means for CAIs that are close together, their broadband accessibility is likely very comparable. We believe this means Broadband accessibility may be less about connectivity than it is about the ability of a CAI to successfully adopt and utilize Broadband to support its mission. Therefore, an important part of where SBDD mapping and planning come together understands what broadband is used for, potential barriers to adoption, and how it is an essential component in a planning region's investment scenario.

Anchor Institution Connectivity

At this point we have focused our CAI survey attention on schools and libraries[19], with respect to connectivity. We benefit from strong relationships throughout the education sector (K-12 and Post-Secondary). We have also found excellent resources within State librarians in all States.

To supplement the education and library information we have formed organizational relationships with the major hospital associations within each state. Our goal with this relationship is to cull information from their planning process. We continue to formalize/advance this relationship. For this data submission, we are also relying on a 3rd party estimate of connectivity needs at these root nodes of healthcare networks--Hospitals.

[18] http://en.wikipedia.org/wiki/Tobler's_first_law_of_geography. We are attaching connectivity based upon the highest speed wireline provider in that block. This provides a ceiling for what can be obtained although the CAI may not be purchasing this level of service based upon needs, budget, mission, etc..

[19] Wisconsin schools and libraries were recently awarded an ARRA connectivity grant < http://dpi.wi.gov/eis/pdf/dpinr2010_27.pdf>

As in the prior submission, we are using public domain sources of information for public safety-category 4. The vast majority of these locations are estimated with respect to connectivity. Our hope is that in subsequent submissions, we will reduce the size of this category and connectivity information specific to root nodes of the public safety network--such as County Emergency Operation Centers.[20]

We have expanded our CAI categories to include museums, convention/visitors bureaus (category 7) and County Seats or Courthouses. Our planning partners believe these two classes of CAIs are a critical engine of economic development and opportunity. They are also very important to supporting civic pride, quality of life, providing access to government services, and attractiveness for location decisions. Our hope is that adding them into the other community support--non-governmental and governmental categories--provides a planning foundation for this class of CAI. At this point, most of the connectivity information is an estimate.

Because we have a wide ranging population of CAIs in our data set we have a wide variety of Broadband services that don't always fit NOFA parameters. Services like PRI or T1 are classified into" other copper", but the bandwidth is estimated based upon the number of channels purchased. We also had difficulty obtaining both the upstream and downstream channel capacities. In large part, we made the speeds symmetrical, but this is an assumption on our part.

Middle Mile

Middle mile information was collected directly from providers via survey or interview. Middle mile is a "chicken or egg" type of challenge in that it is possible to verify that the infrastructure exists, but extremely difficult to know what it is doing without engineering level assistance. Although most providers submitted "something," there was a significant variance in what that "something" represented.

The purpose of this section is to record some of the comments and questions we have received about Middle Mile. We hope this provides better context for our data submission.

The content within the NOFA, Middle Mile was defined as (a) a service provider's network elements (or segments) or (b) between a service provider's network and another provider's network, including the Internet backbone. (Collectively, (a) and (b) are "middle-mile and backbone interconnection points.")[21]

Given the existence of the "or" in this definition, providers submitted a variety of information. Based upon the NOFA example, several fixed wireless providers interpreted Middle Mile in terms of the connection points from their towers to their own serving backhaul location. The topology was

[20] Within the public safety category, it is also very difficult to derive precise locations as many CAI are addressed to PO boxes.

[21] From http://broadbandusa.gov/files/BroadbandMappingNOFA(FederalRegisterVersion).pdf at 54, visited March 28, 2010

commonly microwave from their distribution towers to their NOC . The NOC and towers were listed as the Middle Mile points. This seems to be consistent with the first definition clause (a).

Telephone, Mobile Wireless, and Cable providers tended to remain either silent on the question, or would provide a single location in which Internet peering occurred (clause b). A number of participants explained that the question was quite ambiguous with data traffic moving back and forth over both TDM and IP networks--it was unclear where the distinction should be drawn. As a general rule it seemed like many providers listed a single location where Internet Peering occurred.

A number of providers refused to answer the question on grounds of confidentiality[22]. Others would not disclose as their middle mile points aren't owned--another company provides the physical and electronic connection to their network. In other words, the entity providing Broadband is not the entity providing Middle Mile.

Within a given submission there were two final attributes that tended to confuse. First, should speed be measured in terms of only data capacity and what exactly is "data" (e.g., can/should you segregate out voice or video), and is the relevant capacity that of the physical connection channelized to a specific virtual circuit on their network.

Finally, a number of other providers were unsure of the height above grade measure (is this their floor, the street outside, etc). We seem to have a combination of height above or below grade, as well as heights above mean sea level (AMSL).

To the extent possible in our timeframe, we verified the location of several Middle Mile points. Where we could see infrastructure that appeared to be consistent in location with other provider infrastructure, we felt that the location was accurate. In some cases, the point provided seems sensible (is on a road, near other equipment) but using imagery, we couldn't find a place where this type of connection could occur. This wouldn't be unforeseen, in that much Middle Mile connectivity likely takes place in a protected environment much smaller than a standard Central Office installation.

Mobile Wireless Coverage

We have received mobile wireless coverage from most mobile broadband providers in each state. At this point we have cleaned the geometry of the data and attributed it with spectra and FRN as required.

Provider derived coverage has been reviewed against the commercial licensed product for consistency. We also use licensing locations and tower infrastructure to spot-check supplied coverage. This mode of verification remains complex, given the lack of facility-based information with mobile wireless.

[22] As received in email 9/30/10, "Due to security concerns and the risk of public disclosure of highly sensitive data, whether inadvertent or otherwise, ***REDACT***response to the middle-mile and backbone interconnection request is limited to publicly available information available on {remainder not included}"

Verification

Almost by definition, data verification is an ongoing and evolving process. Clearly with each new data submission there will be a validation process at hand and at the same time, our team continues to expand and improve the efficiency and effectiveness our data verification routines. Consistent with requested changes back to Census 2000 geography and the movement toward an fGDB export database, much of our validation effort was spent in supporting the ETL processes into these formats. These two program changes accounted for more re-work than we had anticipated. In future data submissions we will continue our work to stabilize and improve the business process which normalizes provider submissions into NOFA formats and expand in more depth on the confidence analysis within the data.

Verification Standard

Our overall verification standard is focused on the level at which we supply processed data to NTIA. This means that the vast majority of our verification process will be focused on ascertaining coverage for Census block's less than 2 square miles and covered road segments.

At this stage our working hypothesis (confirmed by our experience) is that there won't be a single dispositive measure to indicate broadband availability in a Census block or along a segment. From prior work, and examining our current provider submissions, we believe that there is too much variation below the submitted record to make a single binary yes/no indication. Rather, there will be a series of measures that combine to provide qualitative confidence (a classification scheme) in our indication of broadband availability at the block, segment, or wireless polygon level. We believe such a qualitative confidence scheme is relevant to/supportive of NTIA interests, as well as the interests of our end-user community–that is, the states and citizens we serve through this program.

The intent of this section is to illustrate why we are moving toward a particular verification methodology. Our team is learning as we go along, and will adjust and improve this thinking, but given our experience to date, this is where we are heading. As stated above:

- First, verification is at the level of data submitted to NTIA.
- Second, verification is enhanced when there is a secondary measure of availability (such as infrastructure presence or serving area boundaries)
- Third, given the limited resources of this effort, the most important verification process to implement is the erroneous dispersion of coverage. These are the "islands" of coverage isolated by significant distance from other covered areas.

Before explaining our overall verification thought process, we have one example to illustrate this very common sub-block level variation type problem.

The case is taken from a gentleman who requested a map change in Alabama. His home is near the yellow dot. The darker grey blocks are covered Census blocks. The black lines are covered road segments. He cannot receive DSL from his incumbent provider although his neighbors can. The incumbent carrier does have at least one structure in that block from which broadband services can be provided; unfortunately his home is not served.

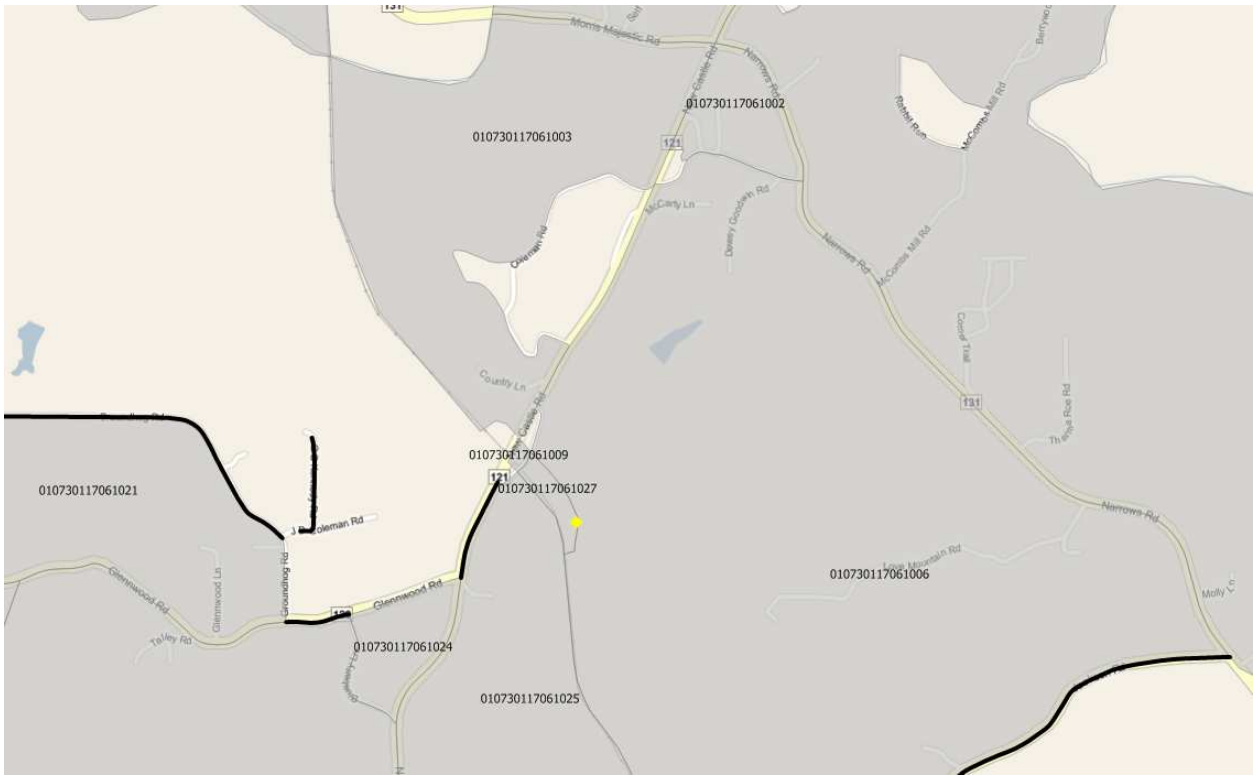


Figure 14--Sub block variation

Because the SBDD program requires the depiction of coverage at the block level, the above map has been correctly generated. However, from the customer's point of view, the map is inaccurate. This requires us to explain that the maps are not intended to be a structure-level qualification, at which point some consumers question the value of the maps when seeking service information. Of course, we also share this information with the incumbent carrier in the area so they are aware of a potential customer market.

Beyond this type of one-off structure-level qualification, sometimes, as shown below, we have even larger gaps in provided coverage. The image here shows an "outlier" block that could be an error, or it could indicate missing blocks along a major road that should have been filled in. In this figure, the outlier block is highlighted in turquoise.

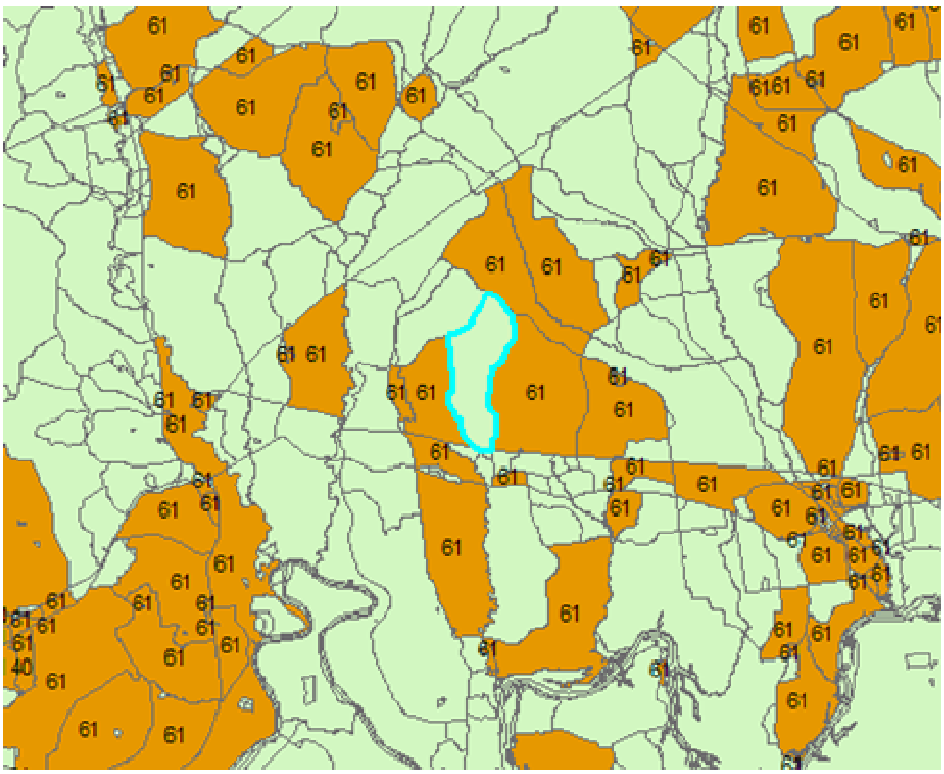


Figure 15--Dispersion in Submitted Data

In this particular case, we are faced with a different verification question. Based upon the properties of the neighbors, we believe this block should likely be covered (coverage interpolation,) but supplied data from the incumbent says otherwise.

The next example, at a somewhat larger scale, shows where an interpolation process requires some adjustment. The figure below shows a town level. There are some smaller blocks that are likely covered by interpolation logic, but we also do not want to extend coverage beyond a franchise boundary as in the areas shown in a box on the bottom of the map.

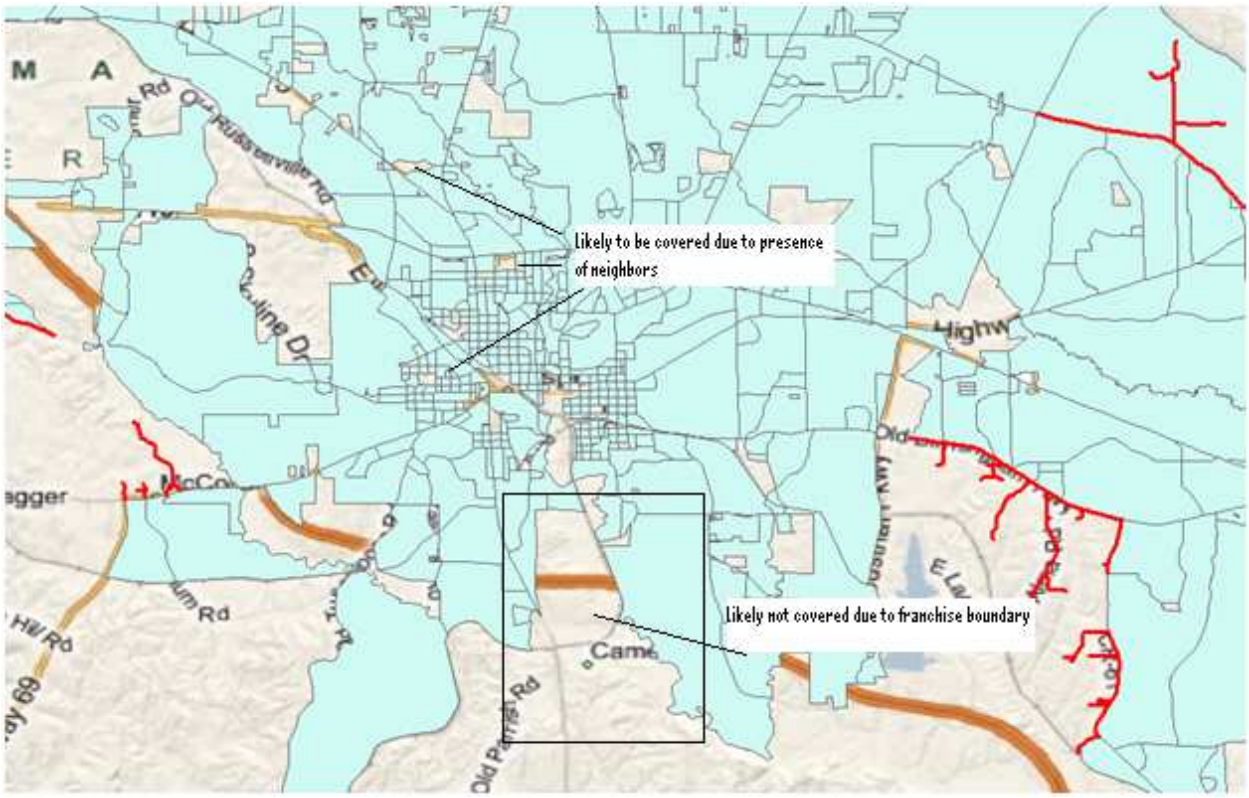


Figure 16-Where do you stop interpolating?

From what we can gather from providers, the submitted data comes from geocoded billing records. In this paradigm, this means where there are no customers; service is not identified on a map. The interpolation verification question then takes on two dimensions.

First, if a provider has no customers in an area, how can we know if they would be able to provide service in a 7-10 day interval?

Second, if we use the properties of neighboring blocks to interpolate coverage, when should we stop (e.g., at a franchise boundary, at a certain distance, etc.)?

We continue to work with providers to get additional information to help us better understand and display this type of circumstance. As of yet, we have not been entirely successful at getting franchise boundaries which would address much of the issue.

The final map shows this dispersion problem, but to yet an even larger degree. This solitary large block is likely the result of a bad geocode, but we don't know, given the data that has been submitted by the provider and the "single customer in a block standard" set by the clarification.

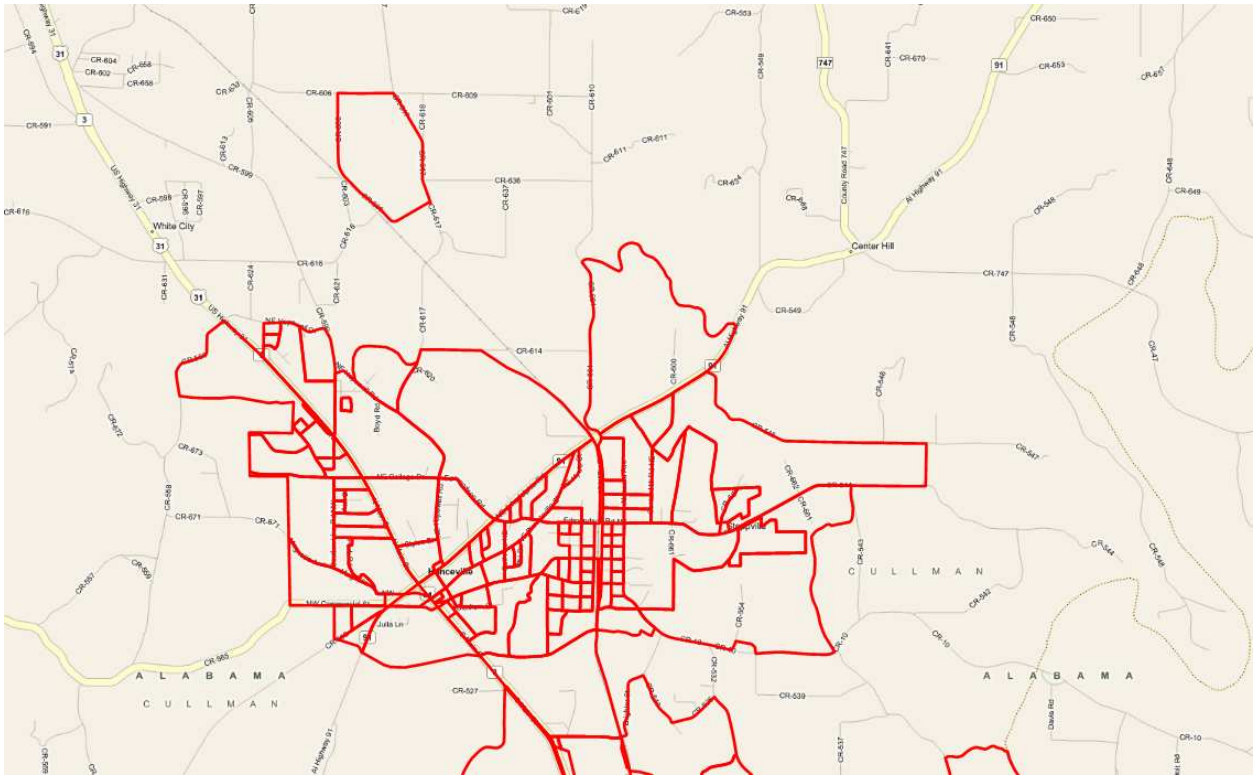


Figure 17-Dispersion in covered blocks

Due to the fact that this situation is quite obvious in display, this type of problem is one that we are more aggressively trying to resolve. Where a single block has no neighbor offering comparable coverage and is a specified distance beyond an exchange boundary, our approach has been to filter these blocks out. As of now this filter is limited to incumbent telco providers because we have a good source of exchange boundaries. The actual query used is based upon the proximity of the isolated blocks from exchange boundaries. Our experience with using a distance measure has been complicated by how to handle segments and segment discontinuity.

In sum, the variability in our source data continues to suggest that our dynamic verification process is relevant and appropriate. And as noted above, we believe the more meaningful outcome of which, will likely be a series of qualitative indicators or expressed confidence levels. Our concern, as with the development of any sort of classification process, is how rigid we should make this classification.

Verification Work Process

To support our dynamic multi-factor verification process, we have implemented the following steps.

First, when data is received, an analyst reviews the submission and as quickly as possible sends back to the provider any immediate questions or concerns. We have found this gatekeeping step very helpful in making sure we understand the intent of the submission.

Second, for all providers who submitted data to us in the first round, they received both a tabular data summary and a mapped output. Prior to releasing the “check maps” to providers, we had a team of

analysts visually inspect each provider's coverage area. The focus on this QC effort has been to identify and flag suspect blocks. After this in-house review, we solicited a second level of feedback from providers and received a number of requested changes and corrections used in the development of the October, 2010 dataset.

For those providers who submit only block or segment level coverage (i.e., in those cases where we have no infrastructure to test with) we test for coverage containment within known service boundaries. The intent of this validation step is to remove blocks that are obviously erroneous.

As mentioned in the sections above we have implemented a check on disperse blocks but we have implemented less with respect to coverage interpolation (holes in coverage). We continue to work on a series of mechanical tools to assist with the inspection process but have run into challenges related to geographic basemap and timing.

As our first submission has moved online, we have also begun to benefit from citizen/viewer and provider feedback. In some cases this has helped us identify and fix errors in our underlying data. In other cases, as we have shared with NTIA, we have encountered some perceptual issues rooted in how the data are developed and modeled to comply with the NOFA. Depiction of uniform coverage in small Census blocks continues to be a challenge.. Despite our best efforts to explain the full block shading requirement, we continue to receive complaints that the coverage shown on the map is not accurate for a particular location within that block.

Online Map Experiences

Now that our maps are online, we continue to harvest viewer feedback and comments. Because an online map allows someone to zoom in far below the scale of the data, a large number of comments reflect sub-census block concerns. While important to the citizens reporting these issues and to our broadband planning teams, this level of data is outside the scope of our core validation process, which as noted above, is focused on the level of data submitted to NTIA.

There are several other themes that our team believes are important to share. These comments are actually quite helpful because they also improve our data processes to better meet the needs of map viewers. For example, we have invested significant time in harvesting more segments from provider data. Because the appearance of segments is so important, we are putting time into ensuring a visually appropriate edge match between the roads we harvest and the blocks/roads we will show online. On a technical level, we also believe that a good segment process will help us understand more about dispersion in the data, and what is valid versus what is not valid.

Perception of Unfair Treatment Across Technologies

Several broadband service providers have expressed concern regarding how wireline services are displayed, as contrasted to how wireless coverage is displayed. This is an artifact of the SBDD data model and the comments that follow may be helpful to NTIA staff as national map alternatives are explored in the months ahead

As an example, consider the figure below.

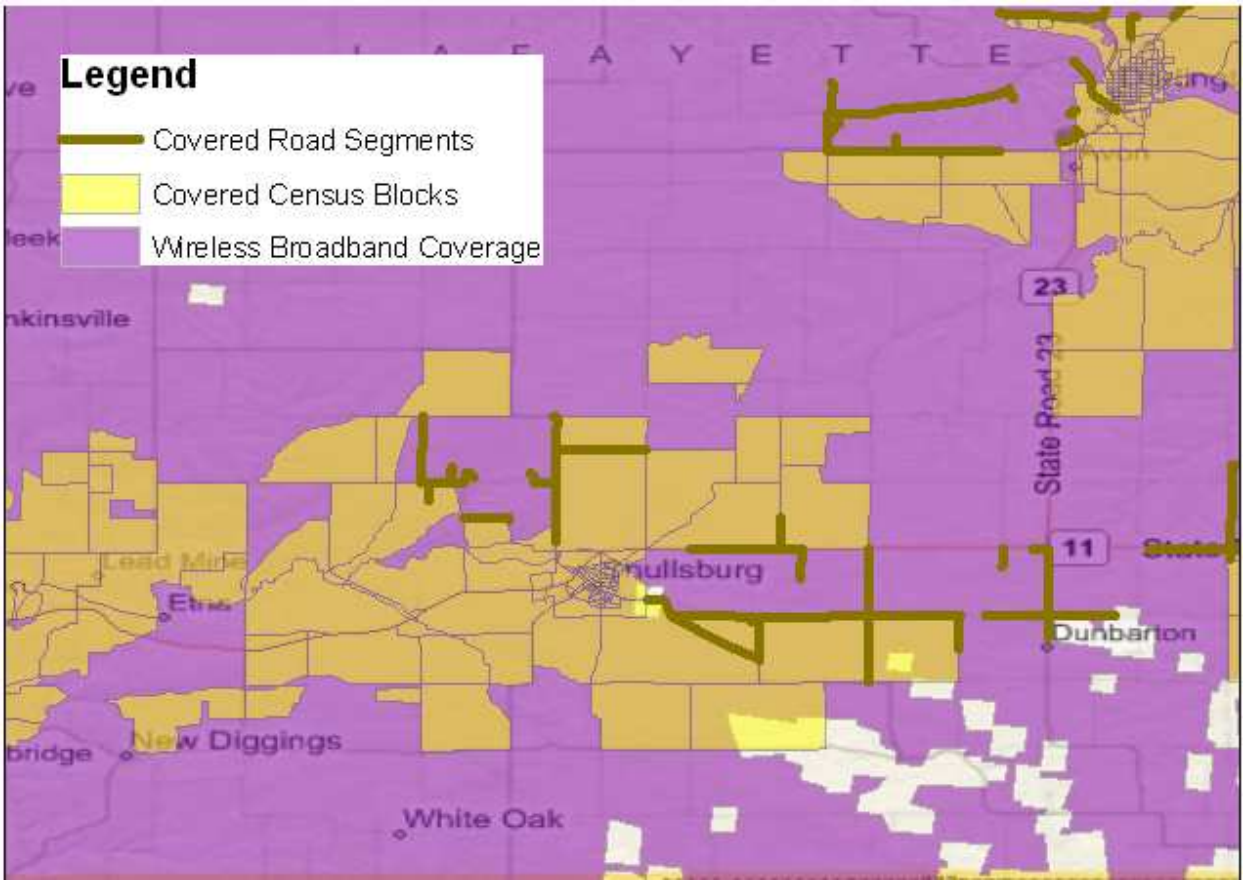


Figure 18--Multi Network Coverage portrayal

In this image covered Census blocks are light gold. Covered road segments are a darker gold and wireless coverage is purple. The concern seems to come down to how a wireline provider's coverage is shown in the large Census blocks (greater than 2.0 sq mi). Wireline providers have expressed dissatisfaction that their coverage is only tied to road geography, which leads to a visual "hole" in their coverage map. At the same time, they feel that it is unfair that the wireless provider's coverage is shown to be uniform in the same area. Put another way if our maps show wireline in terms of blocks and segments, why don't our maps show wireless the same way?

Perceptions of COLR Obligations

Wireline providers have also expressed dissatisfaction that online maps limit the distance of coverage from a road segment. In our current online maps we buffer a wireline carrier's service 300'. A number of providers express that they are mandated to provide voice coverage (which Broadband will accompany) anywhere in the Exchange. There seem to be many dimensions to this argument, but the basic concern comes down to not being able to accurately reflect the scope of their COLR obligation within the mixed block/segment view.

Intentions of Coverage Mapping

When a viewer of an online map clicks on the map (or zooms to an address), they are provided with a pop-up of service provider coverage in the area. The critical question is this: what is the area to which

that pop-up window responds to? Right now, we report back to the Census block, or buffered road segment. As far as the map is concerned, once we move off of that road, or out of that segment we have a new area to examine. Our sense, though, is that our provider view should be a bit more tilted toward finding providers in a general area, rather than finding providers at a single-click location. If the goal of the map is to get someone to call a provider for service, our bias should be to include all of the potential providers in the area, rather than giving potential customers a method to self-disqualify. That is, we want to cast a wider coverage net, rather than one too narrow. The problem with this approach is that it will create a number of false positive broadband reports. We have yet to test this looser standard to see how providers and consumers respond.

Appendix One

Data Collection Challenges

This section summarizes some of the challenges we have experienced with data collection and processing. The team believes it is important to categorize these challenges as they help inform the geoprocessing and verification methods used. It is also our hope that some of the more global issues can be discussed and decided within the Grantee community.

We begin with several global issues and then continue toward more granular challenges.

Global Data Collection Issues

Census Block and Road Standards are not clear

Most carriers submitting Census level information provided 2000 blocks. A few provided 200? or alternative (TeleAtlas) blocks. Especially with the need to derive segment geographies, we would prefer to message to providers a specific Census standard—but we'd like to be consistent with other Grantees so as to minimize work from the provider community. As of now, that standard is Census 2000. If NTIA anticipates using Census 2010 for Spring 2011 collection, it would be helpful to message that as soon as possible.

Also there seem to be several methods by which providers are calculating the area. So the distinction between at 2.00 square miles can be uniform, it would be ideal to articulate an operational area calculation definition as early as possible.

Providers Not Wishing for Block Level Aggregation of Their Data

Both ***REDACT*** have supplied address point level data. Both carriers want NTIA to have the point level information and they request that CostQuest/LinkAMERICA do not aggregate their coverage to blocks. Other than a verification to make sure that point data were contained within, or fell within 1 mile of exchange boundaries, the only other processing was normalization into NTIA formats.

Broadband Providers not Meeting the NOFA "Provider" Definition

PBWorks appears to reflect a concern among a number of grantees about what is a Broadband Provider and given the definition of a Broadband Provider, how that impacts mapping.

If the 7-10 day provisioning rule is to be strictly enforced, it would seem to eliminate a number of prominent broadband providers[23]. Further, the need for clarification around facilities based provider, versus the reseller, has injected even more ambiguity into the mix. Right now we are unclear on how strictly to interpret either of these important distinctions, but we are concerned that we are beginning to create an NTIA exclusion criteria that is going to confuse downstream consumers of the data. In our

[23] By email ***REDACT*** informed us they could not provision in 7-10 days, but they also supply information on qualified locations to the address point level. Therefore we draw a distinction between an incumbent provider owning the facility which terminates at a customer premise who cannot turn up service at a qualified location versus a provider not reporting any specific qualified locations in which they cannot turn-up service in the 7-10 day window. In the first case we have a sense of where service can be offered and verified. In the second we have no evidence that a service could exist there until a specific location becomes a customer.

current submission we have excluded a number of important service providers-mainly geared toward the large business/enterprise market--who don't meet the 7-10 rule. Our preference would be to include them, but qualify their coverage at the Service Overview table level.

Again, we don't want to exclude a service provider, but we believe there needs to be further clarification around the 7-10 day 'rule', reseller definition and interpretation of facilities based provider versus equipping UNEs, SpA or leased lines.

Because of this confusion, there are providers with broadband service data, which we are holding out of the NTIA submission. [24]

To What Extent Should We Begin "Classifying" the Data and Maps?

The question immediately preceding gets to the intent of a Broadband Provider. This question gets to the intent of the Data and Maps.

Earlier in this document we discussed the question of what type of bias we should introduce to our online map messaging. In an online environment, do we want to more likely create an overstatement of coverage for a provider than an understatement? In other words, is the larger problem allowing a consumer to self-disqualify, versus call a number of neighboring providers? There is a related issue to this. Clearly in our maps there is a lot of scatter in data that we believe should be more continuous. These are the islands of coverage from an incumbent provider[25]. There are a number of processes that could be put in place to deal with this type of scatter, but without more information from the service provider-- essentially the last mile facilities-- it will be difficult to perform this clean up in an informed manner. On the one hand, we can aesthetically clean the maps up and reduce the scatter, but we have little sub block engineering information upon which to make this decision. Right now our preference is to put out a somewhat aesthetically messier deliverable and work with providers to get better information to clean their submission. If that isn't forthcoming, we are limited in what can be done given the lack of facility level information. In summary this yields two questions

1. In our online maps should we err on overstating coverageto prevent consumer self-disqualification?
2. In our online maps should we work to clean up a lot of the scatter that we see without having facilities based evidence from which to remove it?

Granular Data Collection Issus

Non-Uniform Submission Standards

It is clear among providers that there isn't a consistent method used to derive broadband coverage. Some providers appear to be using a geocoding approach and then point in polygon or point on segment process. Others may be using GPS locations. In some cases, it is difficult to infer what reference data

[24] Once again, we note that this class of hold out data will influence provider contact and response statistics
[25] For a provider who sells opportunistically (not within a franchise area) it becomes even more problematic to classify their coverage because the points are more related to the type of consumer purchasing the service than a bounded offering.

was used to georeference plant (is it the carrier's roadbase?). This leads to uncertainty regarding the input data scale or accuracy to other base layers. Although we may be trading off absolute accuracy, our standard has been to conflate data to TIGER 2000 blocks and TIGER 2009 roads. We perform our verification against this conflated data product.

Temporal

We are unsure of how well the data are temporally consistent. Some providers gave us their best effort to control to June 2010. We note that some providers were clear that the submission was as of extract date without any way to move back in time. They have no means to control for time and cannot provide any audit support beyond when the data are released to us. Some data-especially loop qualification data-may change from day to day. It will be very difficult to clarify why something was changed from a given point in time.

Perceived Inaccuracy with Respect to Internal Standards

The NOFA is clear on submitting a list of blocks in which a provider delivers broadband service. This is a different objective than perfectly reflecting service territories. If a firm's accuracy standard is a reflection of their service area, than the data created under the NOFA won't meet their perception of accuracy. This leads to two other issues: First, using Census blocks rather than serving area may overstate or understate a particular provider's broadband serving area. This was a significant concern of ***REDACT*** who specifically required us to submit only address-level qualification data. The second issue this brings up, is how or if, there should be some standard on how much of a Census block needs to be covered to call it covered.

Impact of Mergers and Acquisitions

As ***REDACT*** acquisition of ***REDACT*** properties closed during our data submission period, we had difficulty obtaining updated deployment information for former ***REDACT*** territories. As such we derived information based upon the March 2010 submission.

Confidentiality

Several providers have noted concerns with CPNI-related issues and have stated this as a reason for non-participation. We have also heard expressions of comparable concern regarding identifiable responses to Anchor Institution information.

Unclear on Definitions

As discussed earlier, several providers claimed confusion on several key terms involved in Middle Mile. We note a consistent stream of questions around the interpretation of Maximum Advertised Speed. Some providers understand this to be the most common speed package bought within the mass market, others view this as a speed that can be purchased for an additional cost above a mass market offering (eg. a Turbo option for an additional fee per month). Others interpret this as the fastest speed that is available for that particular location--in terms of xDSL a structure qualified speed, for example.

Perception of Data Use

There seems to be some hesitancy releasing speed information because no one is sure how the information will be used as well as what the speed is intended to reflect. A number of providers have

verbally indicated that typical speed will be about (on average) 80% of purchased speed due to overhead. But there are many other factors (such as a user's home network) that influence speeds measures. Providers are concerned about introducing statistics without a clear understanding of how those statistics are derived and then will be used. Also, as advertised speed is pushed down to a block level, we sense more trepidation to report speed values. This quickly begins to touch on parity across network types (why is wireline down at the block when wireless is half the state, etc.).

Location Uncertainty In Source Data

Within this document we have noted concerns about the impact of source data accuracy. Our geoprocessing methodology provided what we believe is a relatively conservative tolerance to account for the scale issue in the source data, but we are unsure of how this may impact downstream users. Clearly, it also impacts the verification process because we can't attempt to verify received data beyond a scale at which it was developed.

Covered Segment Process

Deriving those Broadband covered segments in Census blocks greater than 2 square miles has proved to be a challenge. Moving from a NOFA specified tabular deliverable to an anticipated geographic deliverable also increases the complexity of the effort.

Change Management Process

One thing that is becoming clear is that a change management process that is consistent between the data provider and NTIA is needed. In this light publication of the current data transfer model beyond the PBWorks community would also be helpful . Many providers are designing their data extracts with the NOFA in mind and the NOFA structures have been supplemented in the current model.

Finally, it would be helpful as early in the next cycle to know what Census blocks we are expected to deliver to NTIA. It would also be very helpful to maintain a stable geographic base for the next deliverable so that the basis of verification doesn't change.

Record Level Metadata

It would be helpful to have one or two additional fields in each feature class transmitted to NTIA. One User Defined field could be helpful as an expression of record level confidence. The second field could be used as a Key between the transfer geodatabase and our systems. Ideally both fields could be large text fields (50 char) so the Grantee can use them to express a variety of attributes.

Miscellaneous Data Collection Notes

We note the following important observations regarding our data submission

1. There are middle mile plant records for providers who are not present in the Census block. Segment or wireless area feature classes. This is due to classification as non-NOFA broadband providers.
2. In some cases, we have trimmed wireless coverage estimates to honor state boundaries.
3. We believe some providers are trimming their coverage to honor license area boundaries.

4. Where a provider submitted middle mile points out of state, we retain them and transmit them to NTIA.
5. We had validation errors for BroadbandServed=N records in the CAI table. These records were attributed a Technology of Transfer=0, this cleared validation.
6. In tables with mandatory Zip5 (Service Address) if the End_User_Zipcode was not available, we have inserted '00000'
7. We have a significant amount of VDSL, ADSL 2 and ADSL 2+ coverage categorized into the xADSL category.
8. We have left in the data middle mile locations with above grade elevations that appear to be unreasonable given review of orthoimagery. This seems to be confusion between above grade request and above sea level readings.
9. All fGDB have passed validation except in cases where attributed speeds did not agree with domains associated with technology of transmission (eg Upstream Speed of 2 with ADSL).
10. We note a few providers who have speeds seemingly inconsistent with their technology of transmission. This is either very low speeds wit optical fiber or very high speeds with non DOCSIS 3.0 systems.

Appendix Two

This appendix contains the confidentiality clarification supplied in a series of emails between CostQuest and NTIA.

Feature Class	*Metadata*	*NOFA Confidential?*	*Online Map*	*Public Disclosure*	*Exemption*
Last Mile	Constraints on accessing and using the data	Yes	No	No	None
	Access constraints: None				
	Use constraints:				
	This data is confidential as defined in the NOFA.				
Middle Mile	Constraints on accessing and using the data	Yes	No	No	None
	Access constraints: None				
	Use constraints:				
	This data is confidential as defined in the NOFA.				
Service Address	Constraints on accessing and using the data	No	No	Yes	
	Access constraints: None.				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
CAI	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential

	Access constraints: None.				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
Census Block	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
	Access constraints: None				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
Service Overview	Constraints on accessing and using the data	No	Yes	Yes	The only provider who may not show up this table is a provider who has provided only confidential data (last mile, middle mile,

					address point with provider name)
	Access constraints: None.				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
Road Segment	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
	Access constraints: None.				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
Wireless	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
	Access constraints: None				
	Use constraints:				

	There are no restrictions on distribution of the data by users