

# State Broadband Data and Development Mapping Methodology

---

*For the States of Alabama, Idaho, Wisconsin and Wyoming*  
*Revised March 31, 2011*

*CostQuest Associates*

*LinkAMERICA Alliance*



**LINKAMERICA**

Copyright 2011, All Rights Reserved

## Contents

Overview .....	5
Purpose of This Manual .....	6
Data Sources .....	6
Developing the Provider List .....	6
Provider Outreach.....	7
NDA .....	7
Provider Survey .....	7
Survey Methods .....	8
Follow Up .....	8
No Response Policy .....	8
Summary .....	9
Third Party Data Used .....	9
Confidentiality and the Use of Licensed Materials .....	11
Public Engagement: Crowd Sourcing, Surveys and Social Media .....	11
Consumer Surveys.....	11
Social Media .....	13
Data Production Process .....	14
Data Production Methods.....	15
Deriving Broadband Coverage Information .....	16
Impact of Program Change .....	17
Trends .....	18
Coverage Geoprocessing Methods .....	19
Block Level Coverage Derivation Using Service Point Data .....	19
Block Level Coverage Derivation Using Customer Facing Plant Level Point Data.....	21
Coverage Derivation Using Linear Facilities Data .....	22
Coverage Derivation Using Covered Street Segment Data .....	23
Coverage Derivation Using Serving Area Point Submission Data .....	24
Coverage Derivation Using Polygon/Polyline Serving Areas.....	27
Street Segment Derivation, Large Blocks.....	27
Wireless Coverage Process .....	32

Service Address Point Process .....	34
Coverage Estimation Process .....	35
Speed .....	36
Community Anchor Institutions.....	36
Anchor Institution Survey .....	39
Anchor Institution Trends .....	40
Middle Mile .....	41
Mobile Wireless Coverage .....	42
Verification.....	43
Verification Standard .....	43
Verification Work Process .....	49
Consumer and Provider Responses to Deliverables .....	49
Online Map Experiences .....	50
National Broadband Map Experiences .....	52
Appendix One.....	56
Data Collection Challenges .....	56
Global Data Collection Issues .....	56
Granular Data Collection Issus .....	57
Appendix Two .....	61

## Overview

The following documentation provides an overview of how the third 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.

Although we could separate this draft into state-specific deliverables, the majority of methodology remains intentionally consistent among the states. As one important validation test is comparability across states, we find value in this cross-state approach. This cross-state approach also helps the LinkAMERICA team focus on comparable outcomes across the four states, where appropriate. Our intent is not to make the states look and be the same, rather it is to leverage economies of scope and scale among the business processes.

As expected, this document rests heavily on the prior drafts, but has also been updated and expanded.

Significant changes include additions covering:

1. Trends in provider inputs
2. Expansion in retrieval of WISP coverage
3. Requested modifications based upon NTIA guidance
  - a. Inclusion of satellite, changes to service overview table, FRN verification process
4. Consumer Feedback, Crowd Sourcing and Social Media campaigns.
5. Development and posting of a Technical Standards document.

Treatment of the following subjects has been expanded:

1. Community anchor institutions and survey methodology
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. We continue to focus resources on establishing stable business processes to track submissions, verify received and processed data, test for temporal stability and provide reporting deliverables consistent with NTIA expectations.

In our view, the 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 and modifications 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 processing 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.

We close this methodology document with two Appendices. Appendix One describes 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. Appendix Two describes the confidentiality framework explained by NTIA.

## Purpose of This Manual

This technical document was developed to provide transparency in our data production process.

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 at project inception from the following 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/research efforts
- Interviews with key state staff members and important community influencers

After the October 1, 2011 “Round 2” submission, we continued our research and added new providers to the program as discovered. As one would expect in a dynamic marketplace, provider identification is an ongoing and important component of our work. Mergers and acquisitions, the use of multiple regional DBAs, the lack of any universal identity management attribute, and the generally complex parent-subsidary structure of many telecommunications companies, make provider identification and tracking very challenging.

In early January 2011, we once again initiated an email and telephone outreach campaign to contact all known providers. This is an extremely time consuming process, but it is necessary to ensure that the list of contact persons remains current, and that providers are aware of data request changes and deadlines associated with each round. Where necessary, we execute new NDAs with providers. In “Round 3”, this effort continued on a daily basis until we reached our final data submission deadline on February 18, 2011. After February 18, we continued to work with providers who were not able to meet the deadline. In most cases were able to “crash” our process to accommodate this extra data, but late submissions continue to create inefficiencies and add costs to the overall program. In Round 3 only providers who

responded in the last two weeks of March were excluded from the final dataset. Data from those providers will be updated this summer and included in our Round 4 submission.

Once again, as contact is made in each round, we verbally qualify each provider by asking a series of questions regarding the type of service and speeds offered. If the provider does not meet the minimum specifications for a Broadband provider (as defined in the NOFA) we make a note of their status and remove them from the data submitted to NTIA.<sup>1</sup> We continue to reach out to them in future rounds in the event that their service is upgraded or expanded.

### **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. As described above, individual e-mails and/or telephone calls were made to all providers explaining the status of the program and requesting their continued support in Round Three. We've also had the opportunity to support providers in their BTOP / BIP applications in certain cases. Through these collective outreach initiatives, and our engagement with various industry associations, 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 1 or 2, they were given an additional opportunity to do so in Round 3. 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**

Since two prior rounds of data collection had been completed, the LinkAMERICA team had a solid base of coverage and speed information with which to begin Round 3. This allowed us to provide two response options to providers. The first was for them to review PDF check maps of their coverage and speed data – submitting only corrections and additions to the existing dataset. The second was to allow submittal of completely new datasets, either in tabular form or in multiple other digital formats. For those without sophisticated CAD or GIS systems, we continued to allow the submittal of printed/scanned maps and other written materials.

---

<sup>1</sup> As with other Grantees, we struggle with appropriate and consistent classification for service providers like Megapath, New Edge Networks, American Fiber. These providers seem to resell and/or provision within their own network opportunistically. In this submission we begin to bring them into the analysis as a provider type "other". As the inclusion of this category isn't our primary goal, we are working to process data as we can. We are similarly categorizing and retaining reseller information. Our datapackage.xls illustrates the categorization of non Broadband providers within our provider tracking and verification systems.

## Survey Methods

Once again, we used a secure digital survey process (via our provider portal websites) to collect and display information for providers. The Round 3 survey process 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 Rounds 1 & 2, if providers could not supply information in the requested format, 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 coverage features for each provider.

Returning Providers: While many Broadband providers submitted datasets in Rounds 1 & 2, many of those submissions did not contain 100% of the requested data. To help identify gaps, and to make the Round 3 submission process as simple as possible, every Round 2 survey was reviewed for completeness, as well as accuracy and formatting compliance. Notes were made regarding gaps, and specific instructions were developed for providers in Round 3. These instructions not only explained what data was missing, but also provided directions on how to include that information in the Round 3 submission.

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

## Follow Up

After the release of the Round 3 survey in early January 2011, 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 January and February. The initial data submission deadline was set for February 18, but, as previously noted, we continued to accept “straggler” submissions well into March.

## 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 valid phone number, no contact person identified) they were removed from the list of “known providers”. We believe the vast majority of those we were unable to reach were small wireless providers who have simply ceased to exist<sup>2</sup>.

---

<sup>2</sup>The complete list of known providers and important submission statistics are contained in the datapackage.xls file.



## Summary

In summary, an intensive 45-60 day provider outreach and data collection process is initiated at the beginning of each round. In Round 3, given the data vintage of December 31, 2010, we began this process immediately after the New Year. The last submissions were accepted in mid-March, 2011.

While we continue to successfully engage the majority of providers in each round, the amount of manpower required to solicit complete and timely responses should not be underestimated. This process is one of the most costly and complex within the entire SBDD program.

## Third Party Data Used

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

- American Roamer, Coverage Right Advanced Services. This data served two purposes. The first was to verify the provider list and help find Broadband service providers not on other lists. The second was to verify the reasonableness of the Broadband service provider's submission.
- MapInfo ExchangeInfo, Professional. This data was used in the verification of telephone Broadband provider data. Where a public domain exchange boundary wasn't available, the MapInfo boundary was used for coverage containment tests.
- Media Prints Cable boundaries. This data was used in the verification of Cable/HFC Broadband provider data. It was used to research valid providers and discover if that provider was offering Internet service. In very rough terms the contained boundaries were used to test the location of some provider data.
- GeoResults Telecom Research Data. This data was used to help estimate the Broadband services likely provided to certain classes of Community Anchor Institutions (CAI).

We have included third party data sources, which touch on each of the three major technologies analyzed within the SBDD program. Each of these data sources tie 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 available, 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 of how we may be able to 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 not?). 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.<sup>3</sup>

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

---

<sup>3</sup> We also suggested forming a technical standards committee and a consistent system for confidence reporting.

- a. Geographic Data Files
  - i. US Census TIGER data<sup>4</sup>
- b. Sources that helped isolate providers, identity management or provider service areas
  - i. NECA Tariff 4
  - ii. State produced exchange boundaries
  - iii. Carrier produced wirecenter boundaries
  - iv. FCC 477 provider filers
  - v. FCC Coals reports (321/325)
  - vi. FCC FRN API lookup tool
  - vii. FCC/FAA Antenna Registration System
  - viii. FCC FRN Lookup Tool (plain text search)
  - ix. USAC High Cost FCC Filing Appendices
- c. Sources that helped isolate anchor institutions
  - i. USAC Grant lookup tool
  - ii. USAC High-Cost FCC Filing Appendices
  - iii. HRSA data warehouse
  - iv. NCES data lookup
  - v. State managed lists of schools (K-12), post-secondary institutions and libraries  
List of museums, conventions, and visitors bureaus from [www.onlineatlas.us](http://www.onlineatlas.us)

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 those specified in the NOFA, perhaps making them less reliable than information collected directly from providers. At the very minimum, provider data has a lineage and temporal status that we can identify. A concern we have with increasing use of third party data is that we have no way to verify its quality or development methodology. In other words, we may hit a wall in which we can't determine how the commercial source derived its coverage conclusion. To us this means that third party data sources are beneficial, but represent a supplementary view, not an authoritative one, of the NOFA defined Broadband market.

In short, 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.

---

<sup>4</sup> 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.

## 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 about 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.

For instance, some mobile wireless providers were unable to submit coverage information to us. In these circumstances we have generalized the American Roamer coverage. For incumbent telephone providers we have used commercial wirecenter boundary products to filter Census Blocks that 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 actual 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.

## Public Engagement: Crowd Sourcing, Surveys and Social Media

Crowd sourcing (i.e., an intentional and carefully designed effort to tap into the collective intelligence of the public at large to expand our knowledge base) continues to be an important element of our data collection and validation process. In addition to the various opportunities, the public has to provide input via the online service coverage maps and the related 'Broadband story' process, our crowd sourcing efforts are grounded in a fairly traditional telephone survey approach, focused on the consumer market. In addition, we are currently advancing our crowd sourcing process to include certain initiatives centered in two social media outlets – Facebook and Twitter. These initiatives are summarized below.

### Consumer Surveys

Working under contract for the state of Alabama in 2009, our initial consumer survey was performed before the NTIA SBDDP grant was in place. Subsequent consumer surveys funded by the SBDDP grant were hosted in 2010 for the states of Idaho, Wisconsin and Wyoming. These surveys will be repeated after two years to establish and evaluate trends. These primarily telephone based surveys include two distinct and carefully scripted tracks: one for internet users and one for non-users. The telephone survey approach allows us to reach the non-internet user group as well as the current internet user. A secondary online approach is also used to augment input from current internet users. For non-users, the surveys help determine why they don't have or don't use Broadband. For current Broadband users, the survey helps determine the nature of their Broadband access and how they use that connectivity in their

daily lives. In addition to our state-specific surveys a nation-wide survey was also hosted to provide a broader view of consumer views for comparison purposes. State-specific surveys are, where possible, framed to match the state's regional Broadband planning structure (e.g., the consumer survey in Wyoming was designed to produce results relevant to the state's seven Broadband planning regions).

The resulting data is helpful on a number of fronts in the SBDDP's mission to advance the access and adoption to Broadband. Survey data provides an important, albeit broad, gauge for assessing coverage information obtained by providers. For example, areas with widely available coverage (according to provider information), but lower consumer subscription levels (according to survey results), or perhaps where survey results suggest Broadband is not available, can be examined in more detail. Survey results are also very important to the Broadband planning (and capacity building) components of the SBDDP program in that they help inform and formulate Broadband advancement priorities. Survey results also help inform Broadband policy discussions on both the local and state levels. Finally, survey results provide important information to the service provider community regarding market demand and specific internet use in specific communities (i.e., regions).

The 2010 surveys were launched in July 2010 with a test number of survey calls to confirm (and adjust as needed) the structure of the survey and the underlying survey process. The surveys were closed on November 30, 2010. Telephone surveys were completely random beginning with the acquisition of a list of state-specific, randomly selected landline telephone numbers (e.g., 80,000 random Wyoming residence telephone numbers were acquired as the foundation for the Wyoming survey). Mobile phones were not included in the initial surveys. Upon evaluation of the survey statistics, an auxiliary survey was executed to ensure younger groups (i.e., age 18 – 25) were adequately represented. This secondary step is required because of the continued migration (by younger markets) to non-landline based communications. This younger market (age 18 – 25) was surveyed by reaching out through social media outlets to encourage their participation in an online survey process.

Survey statistics point to the complexity of the telephone-based survey process. Survey volume achieved statistical validity ranging from a 95% confidence level and a  $\pm 1.7\%$  margin of error for the statewide data in Wisconsin to a 95% confidence level and a  $\pm 3\%$  margin of error for Wyoming's statewide data. Most regions in the 3 states have a 95% confidence level with a  $\pm 5\%$  margin of error.

Call volume and disposition is summarized in the chart below

## BROADBAND MARKET RESEARCH - ID, WI, WY - FALL 2010

	TOTAL		IDAHO		WISCONSIN		WYOMING	
TOTAL RECORDS CALLED & % OF STUDY	106,592	100%	22,144	100%	57,445	100%	27,004	100%
NO ANSWER	53,507	50%	11,974	54%	25,886	45%	15,647	58%
TOTAL DEAD NUMBERS	23,962	22%	4,529	20%	14,611	25%	4,822	18%
HARD REFUSALS	9,304	9%	1,728	8%	6,048	11%	1,528	6%
QUALIFIED REFUSAL	643	1%	101	0%	403	1%	139	1%
BUSY	3,652	3%	754	3%	1,903	3%	995	4%
ANSWERING MACHINE	6,385	6%	1,314	6%	3,388	6%	1,683	6%
NON-WORKING NUMBER	5,072	5%	943	4%	2,983	5%	1,147	4%
CLAIMS PREVIOUS INTERVIEW	113	0%	16	0%	68	0%	29	0%
NON-RESIDENTIAL	454	0%	104	0%	239	0%	110	0%
LANGUAGE BARRIER	1,003	1%	223	1%	562	1%	218	1%
OTHER PHONE PROBLEMS - FAX/MODEM	907	1%	205	1%	500	1%	202	1%
PORTED NUMBER	272	0%	68	0%	149	0%	54	0%
BREAK OFF - SCREENER	556	1%	103	0%	301	1%	153	1%
TERM Q3 - UNDER 18	122	0%	22	0%	65	0%	36	0%
		99%		100%		99%		99%
<b>TOTAL COMPLETES</b>	5,758	5%	1,080	5%	3,420	6%	1,259	5%
<b>AVG Completion Time (minutes)</b>	<b>16</b>		<b>15.8</b>		<b>15.4</b>		<b>16.1</b>	

As noted above, the telephone survey process represented in the statistics above was augmented by providing online access to the survey. Participation in the online survey was promoted on all of our state-specific public web sites and selected social media.

As a final relevant point with respect to the consumer survey process the length of the survey is noteworthy. By survey standards, this was a long survey. As noted above, the survey averaged sixteen minutes across the three states. While this clearly contributed to the number of survey call attempts that were required to reach the level of statistical validity, it was not insurmountable.

### Social Media

The phenomenon of social media is widely documented and yet still emerging as an effective access point for public engagement. We continue to explore appropriate ways to use a variety of social media venues in our SBDDP efforts. All of our efforts are informed by and consistent with relevant state statutes and guidelines. Different states have different perspectives on if and how the state will participate in the use of social media. Some state requirements are well defined and some are still being formed. Where appropriate, we use YouTube, LinkedIn, Facebook and Twitter to support our work. YouTube and LinkedIn postings are used to promote awareness. As noted above, we were able to promote additional input on the consumer surveys through a social media outreach program aimed at our younger market segments.

In addition, we are currently engaged in two specific social media tests (in Alabama) to gauge how Facebook and Twitter can be used to drive public input on two important crowd sourced issues: online speed tests and input on map accuracy. Based on data obtained through our web site traffic monitoring process and readily available social media tracking processes, our most recent results are promising. For example, with a fairly limited 'following' a single Facebook post aimed at driving traffic to the online speed test, had 282 impressions (i.e., the number of times the post was viewed), which contributed to an increase in 71 more visits to the Facebook page generally, and a volume of 60 hits (over a three day

period) on the web site page that hosts the speed test. Our normal volume of speed test page hits is in the neighborhood of 7 or 8 per day (vs. the average of 20 per day experienced during this test). Preliminary data suggests that about half these page hits resulted in a speed test being executed.

## **Data Production Process**

To support our objective of transitioning the data development process to our State partners, we continue to model and document our data production process. We find this to be a very beneficial step for two purposes.

First, it helps us understand why (and if) a task is being done, and if it is being done efficiently. Much of this program started so quickly that it was difficult to plan logical integration and hand off points among the various workgroups. Further, we are currently in the process of consolidating much of the process data (check-ins, check-outs, metadata) and we can use this process model to efficiently plan a cohesive information architecture.

Second, our process documentation and modeling helps explain why resources are being consumed in a particular way. This helps our State partners plan for in-sourcing specific tasks as their time and budgetary constraints allow. It also helps our LinkAMERICA team better plan and cross-train members to deal with the work surge that occurs 30-45 days prior to submission.

Finally, documenting and modeling our process helps us take advantage of increasing specialization and proficiency with certain types of data and management responsibilities. In this submission, we had identified data “czars” responsible for check-in and check-out of data. That data czar helped to bridge the gap among receipt functions, provider feedback, production and DBA.

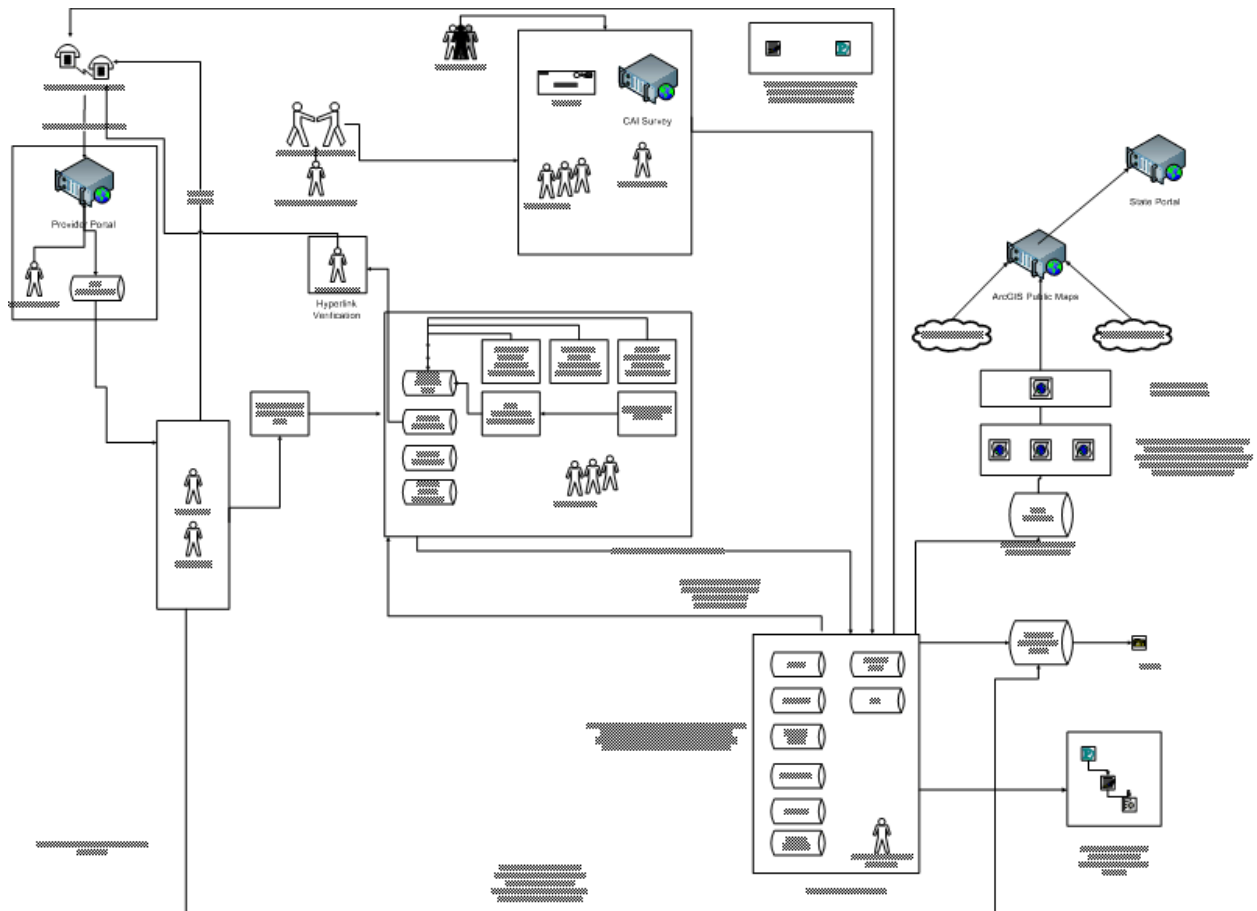


Figure 1--SBDD Business Process Diagram

## Data Production Methods

As raw data were received from the provider community, attention turned to normalizing the disparate submission formats<sup>5</sup>. 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
- Speed certainty
- Temporal certainty
- Provider and network ownership certainty

<sup>5</sup> 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.

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<sup>6</sup> 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<sup>7</sup>
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 –understanding the data and assigning 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. Apply appropriate source metadata<sup>8</sup>

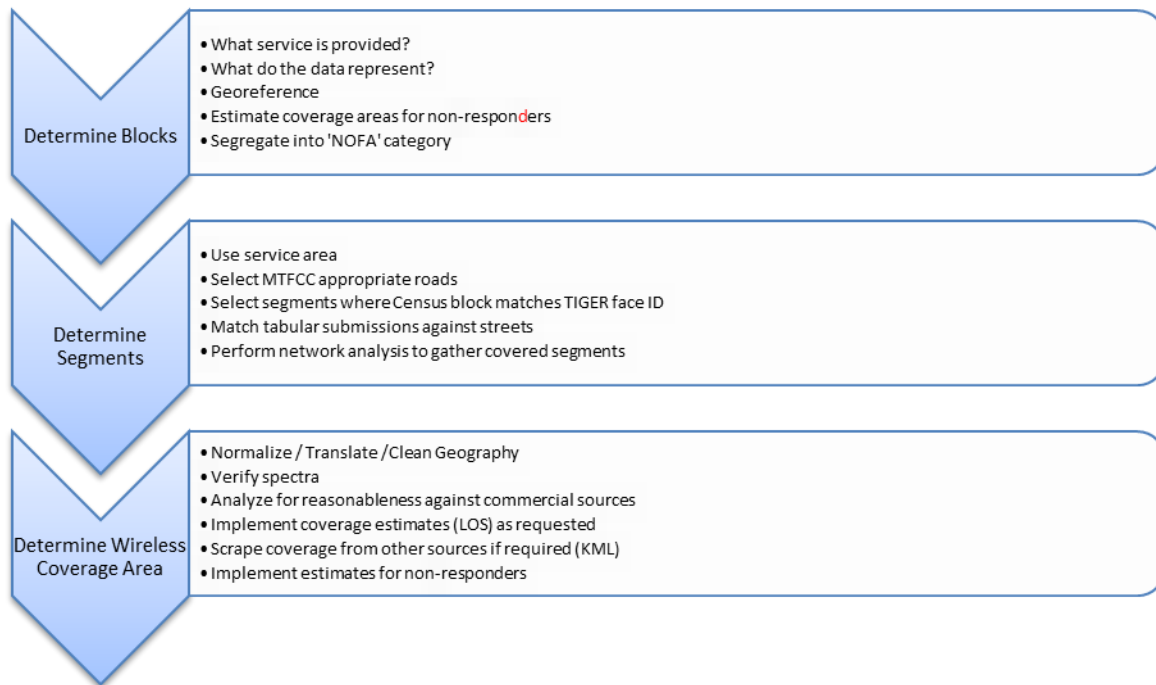
---

<sup>6</sup> Speed, Anchor institutions and Middle Mile facilities are discussed in later sections.

<sup>7</sup> To help clarify issues relating to Census block area and vintages in use, our team [published](#) a technical paper to the Grantee workspace. Because we were unsure if this standard should be implemented uniformly, this document was never distributed to the provider community.

<sup>8</sup> When our team logs a submission into the staging database we record at least two attributes. One records the method used to derive the coverage, the other records the method by which speed was attributed to that object. Other attributes carried to NTIA carry source meta values as well.





**Figure 2-Broadband Coverage Process**

### Impact of Program Change

There were four important program changes that impacted how Broadband coverage was developed and submitted to NTIA in Round 3.

The first was the development of a “provider match” submission metric whereby the grantee’s complete list of known providers in the state is compared against lists from third party sources. The provider match specification was discussed on a webinar prior to the release of the national map. Although, to this date, there has been no clarification on how this metric is established or exactly how it will be used. We have invested significant resources to support an internal process to compare our provider lists with several additional sources. This has been manifest in at least three ways.

Within our provider verification process we work to derive a state level match against third party data sources. As discussed in the early pages of this manual, there is no guarantee that a third party data source is any more accurate than submitted data, nor does it necessarily reflect the provider ecosystem specified in the NOFA, Technical Appendix A. We devote significant resources to matching our submitted data against three, third party data sources. In many cases this becomes a judgment call trying to match provider names across systems. It is a difficult and somewhat arbitrary process. Nonetheless we do believe it has value because it forces a re-examination of who we believe is an appropriate provider within a non-NOFA context.

The use of a provider match system, as well as the webinar comments (3/17/11) directing grantees to estimate, wherever possible, non-participating providers have made us back away from one of our fundamental assumptions in data collection. As discussed in the prior draft of this manual, we had developed a certain “hold-out” class of data when a provider’s data wasn’t of sufficient quality to verify,

or we were unable to put it into the data model (eg. address points submitted for a wireless). In this submission, much of this hold-out data has been included. In some cases this means we are using simple polygons to capture a wireless ISPs serving area. Other times, if we are confident in the coverage, but can get little clarification on the submitted speeds or frequencies, we release the coverage and note in our internal metadata the source issues with the other attributes.

Finally, we have used the new provider type classification of ‘other’ to bring some aspect of the provider’s data into our submission. There still seems to be confusion on how to handle provider types where a provider offers multiple paths to receiving Broadband for typically business customers. Rather than waiting for certainty on the answer, we bring the provider in and list them as Provider Type “other”. Our sense is Provider Type “other” will continue to expand in the fourth submission as we pull in more providers who are facilities-based and reseller.

Clearly one challenge is the data, but an equally significant challenge is appropriate messaging around this “other” provider type category. We do not want to leave consumers with the impression that they can get a high capacity fiber or Microwave link despite the fact that the hospital next to them in the same Census block can get this service.

The final set of changes was a second verification check against reported FRNs. As NTIA is stressing the importance of this attribute, we increased its visibility in our Check Map process. FRN is now listed on both the tabular verification report and the provider PDF map. Beyond this increased visibility we had an analyst verify each FRN in our system against the FCC API<sup>9</sup>, as well as FCC textual search<sup>10</sup>. Because the FRN is not an identity management tool, we are unsure if the FRNs we’ve included are those desired by NTIA, but we have at the very least, verified the existence of the FRN via the FCC system.

### **Trends in Provider Supplied Data**

With this third submission we take note of three important trends.

First, with larger providers, we are seeing an increase in data stability relative to earlier submissions. In informal discussions, several providers have noted changes and stabilization in internal data processes. The firms have invested internal resources in stabilizing this data feed.

We see this reflected in very stable counts of Census Blocks and road segments. This does not mean that complex problems like segment identification or dispersion in data have been ‘fixed’. It does mean that the format and methods to produce inputs for NTIA are increasingly stable.

Second we note that several providers have been particularly concerned with an appropriate identification of Maximum Advertised speeds. In some cases this involves identification of very small areas (sometimes below the level of a Census block) and appropriate assignment to technology of transmission and maximum advertised speed tiers. In other cases, questions arise regarding maximum advertised speeds that could be sold based upon network design, but that are not generally “advertised” or otherwise stated to the general public.

---

<sup>9</sup> <http://reboot.fcc.gov/developer/frn-conversions-api>

<sup>10</sup> <https://fjallfoss.fcc.gov/coresWeb/simpleSearch.do>

Third when comparing submission three results relative to submission two it is important to recall the inclusion of much new data within the Provider Type “other” category. This change does not necessarily reflect a change in the size of the market, rather it reflects new data coming into the analysis and segregated into a distinct category..

### **Coverage Geoprocessing Methods**

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

In most cases, in Round 3 we were still not provided with 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.<sup>11</sup>

### **Block Level Coverage Derivation Using Service Point Data**

A number of providers submitted point level customer data.

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, one of the following scenarios has occurred: 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.

---

<sup>11</sup> 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 that we will inconsistently message this with grantees in other states. We would appreciate consistent guidance from FCC/NTIA on this topic.

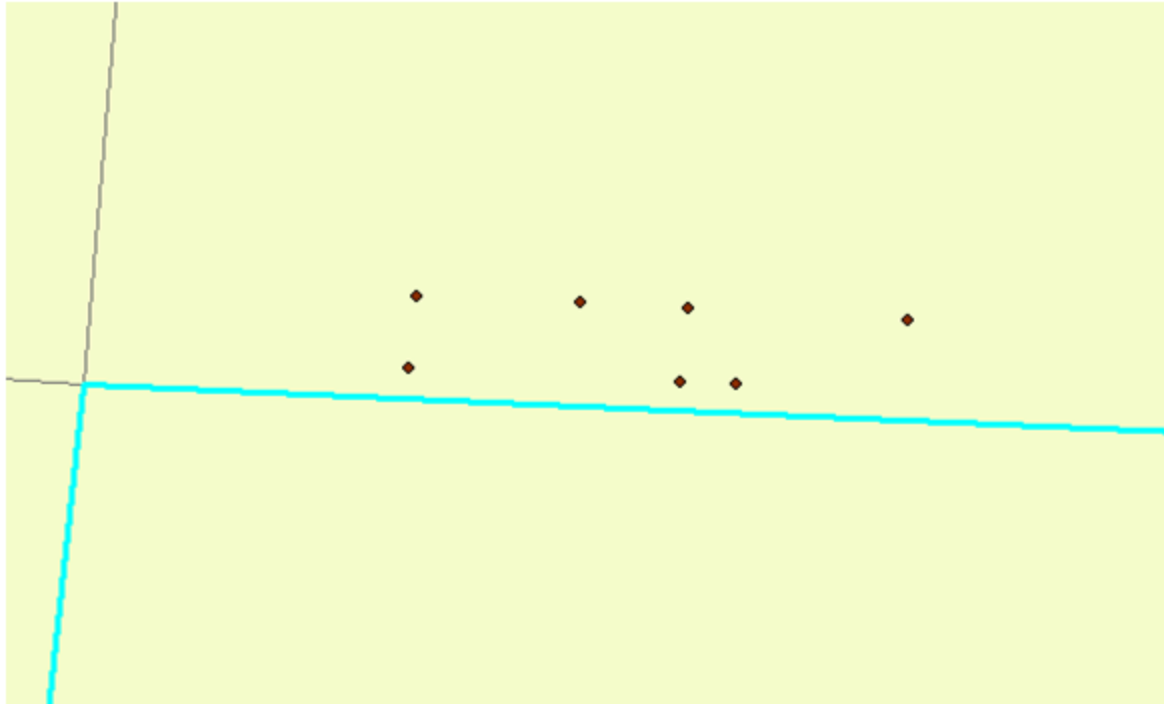
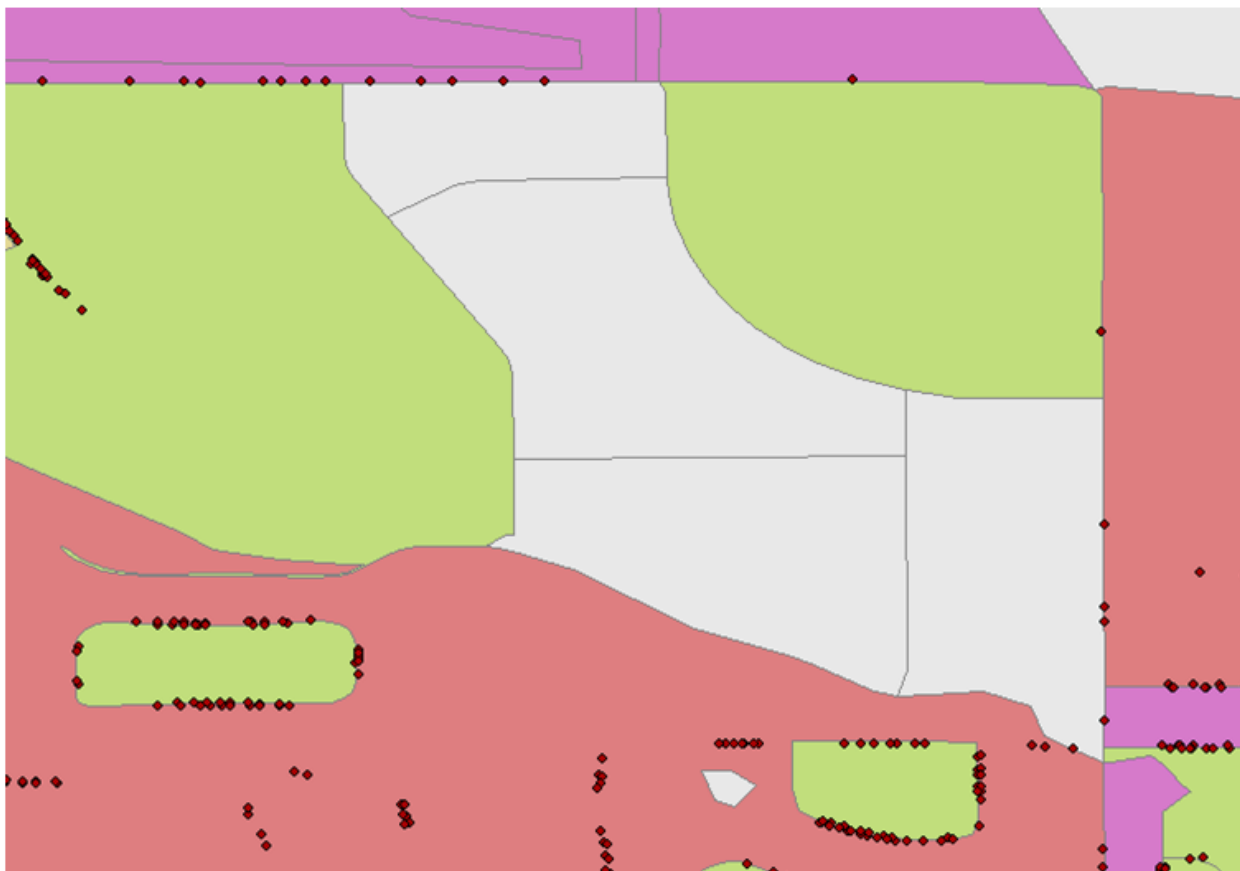


Figure 3-Internal inconsistency in submitted data

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 4-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, 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-foot intersection buffer.

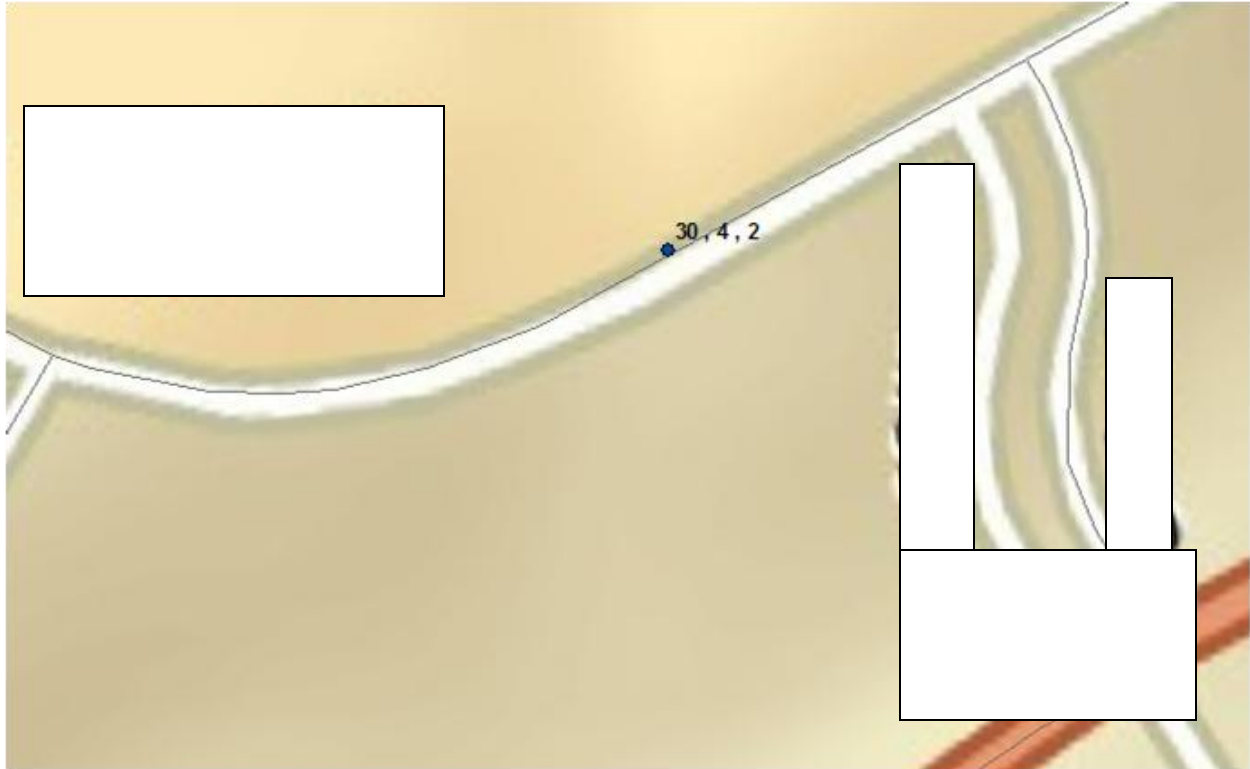


Figure 5-Plant Point level data

#### 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 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.

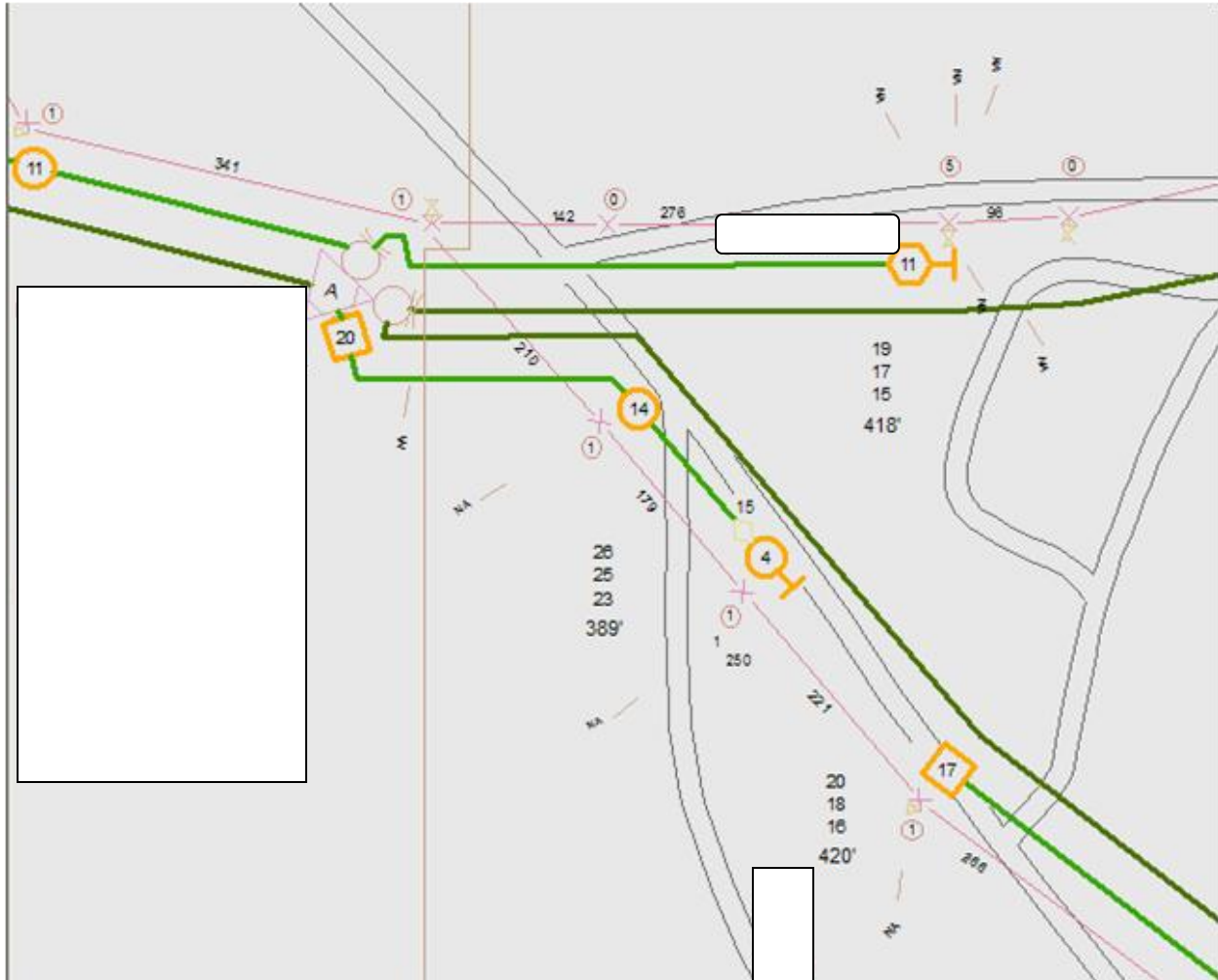


Figure 6-Georeferenced CAD information supplied by Broadband provider

In other circumstances, we were provided engineering information that 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 2,000 meter Census block intersection. The 2,000 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.

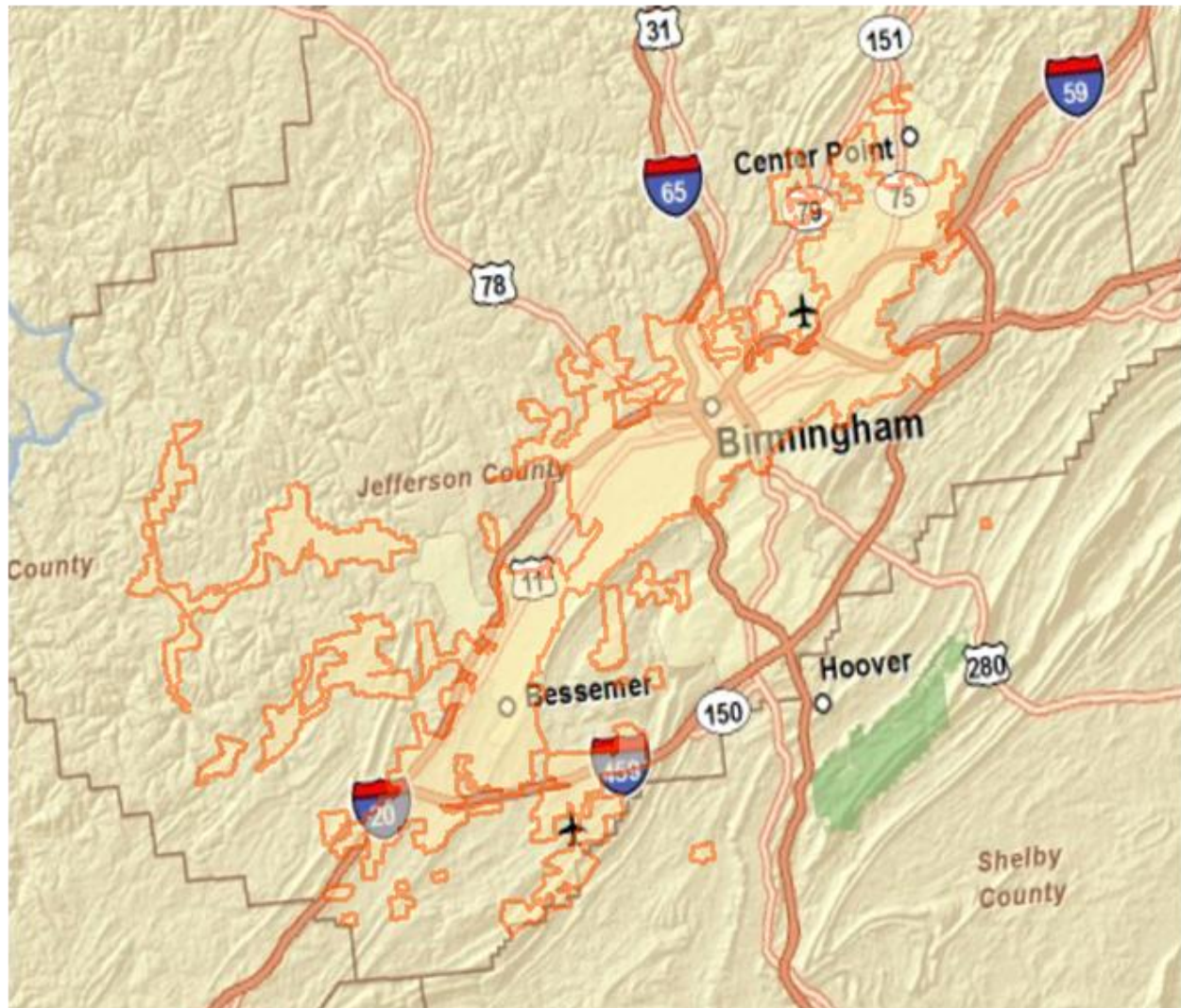
#### Coverage Derivation Using Covered Street Segment Data

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 7-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.

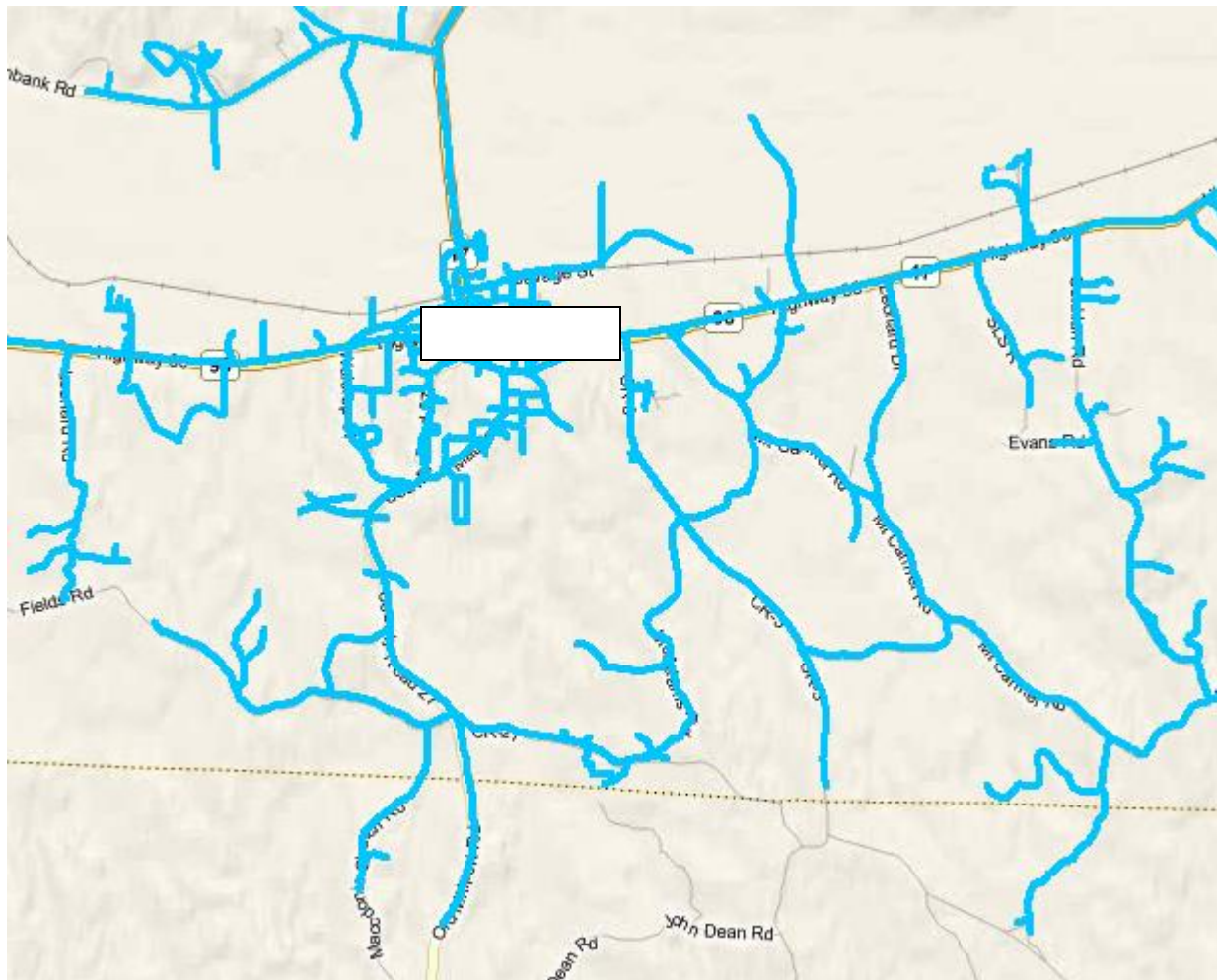
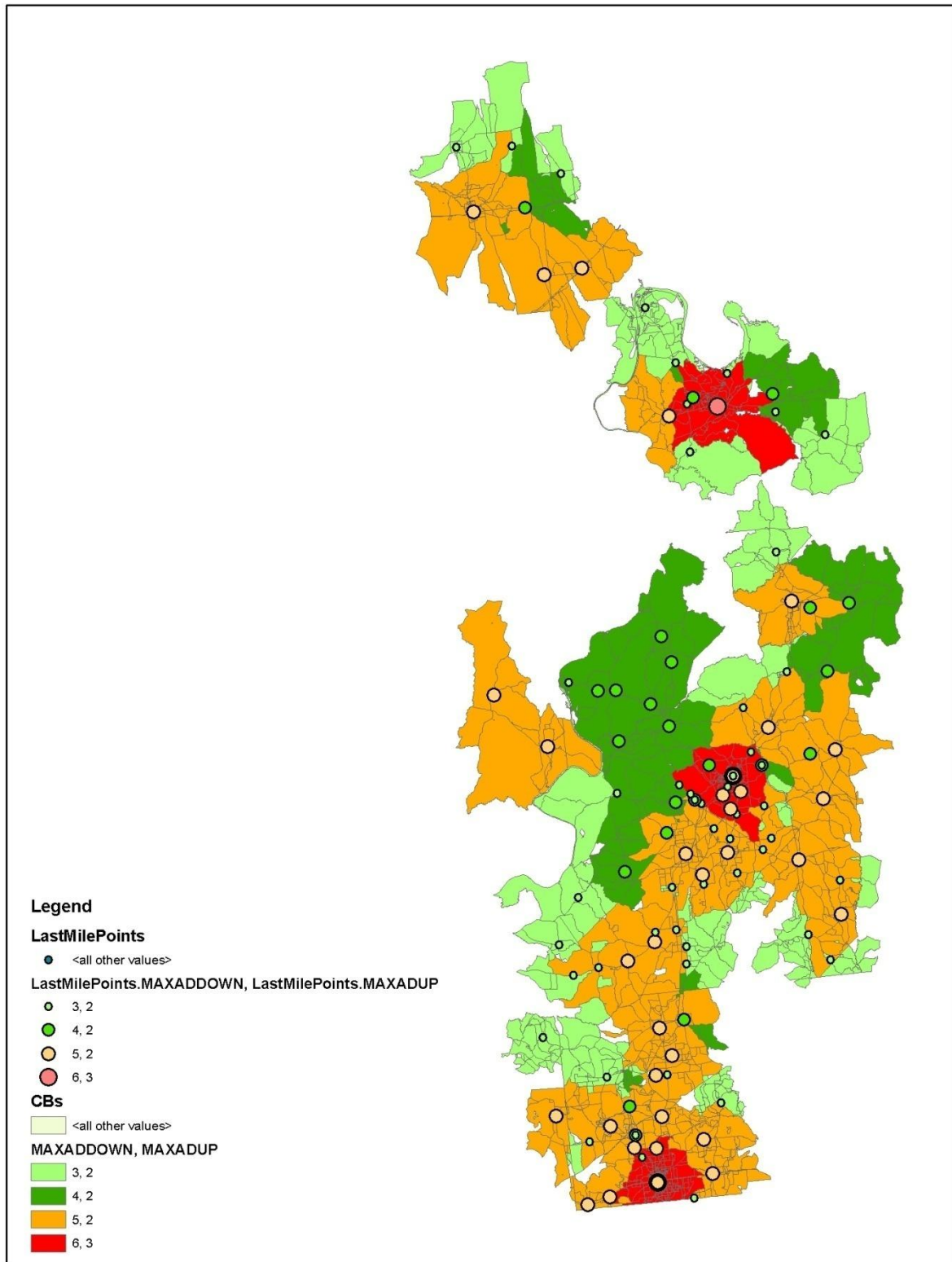


Figure 8-Coverage derived through road paths

In response to Provider feedback we revised this process to include a larger variety of TIGER road types. In Round 1, unimproved roads were not used. In Rounds 2 and 3 -- particularly to improve estimates in areas bordering parks and public lands -- a wider class of TIGER roads was used.<sup>12</sup>

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.

<sup>12</sup>Only TIGER features of MTFCC type S1100 and S1200 were excluded from use.



### 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 did not get picked up within a spatial query).

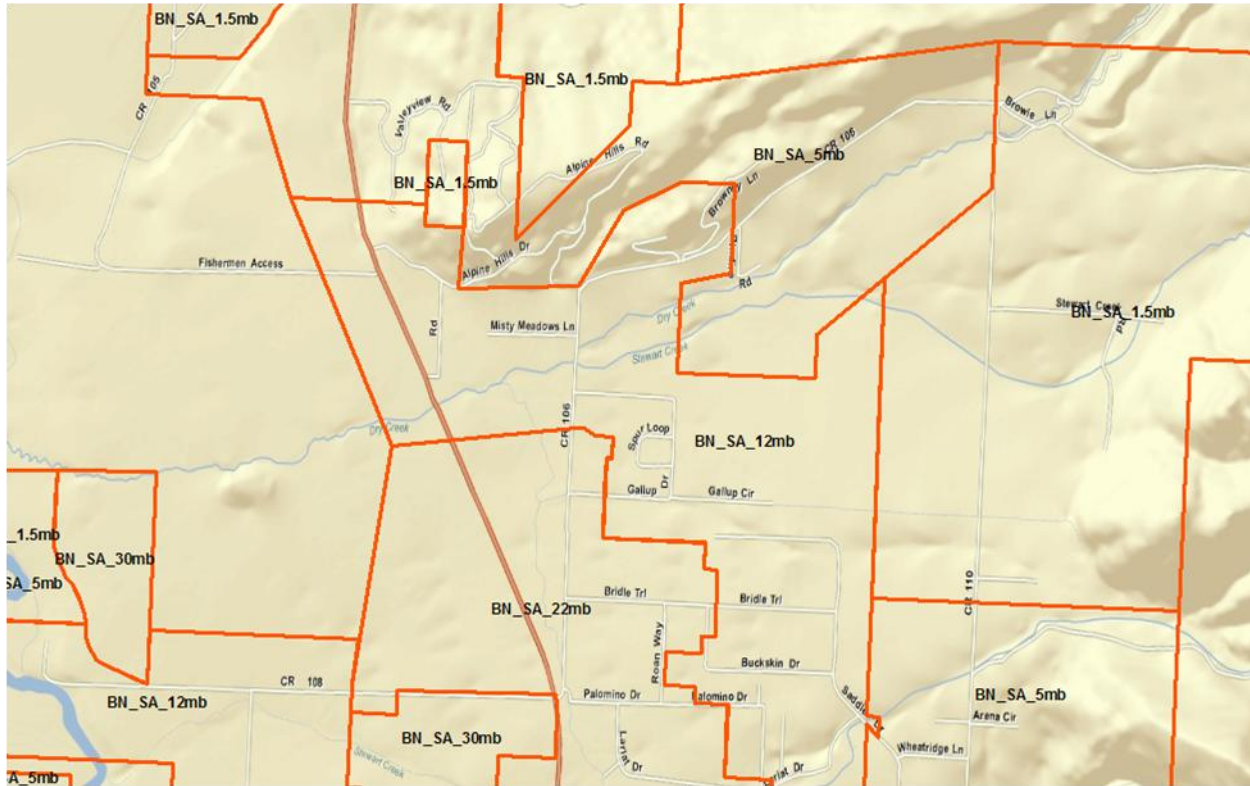


Figure 9-Coverage derived through serving area polygons

Although we cannot directly solve the loss of speed granularity due to Block shapes, we honor 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 30% 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.<sup>13</sup>

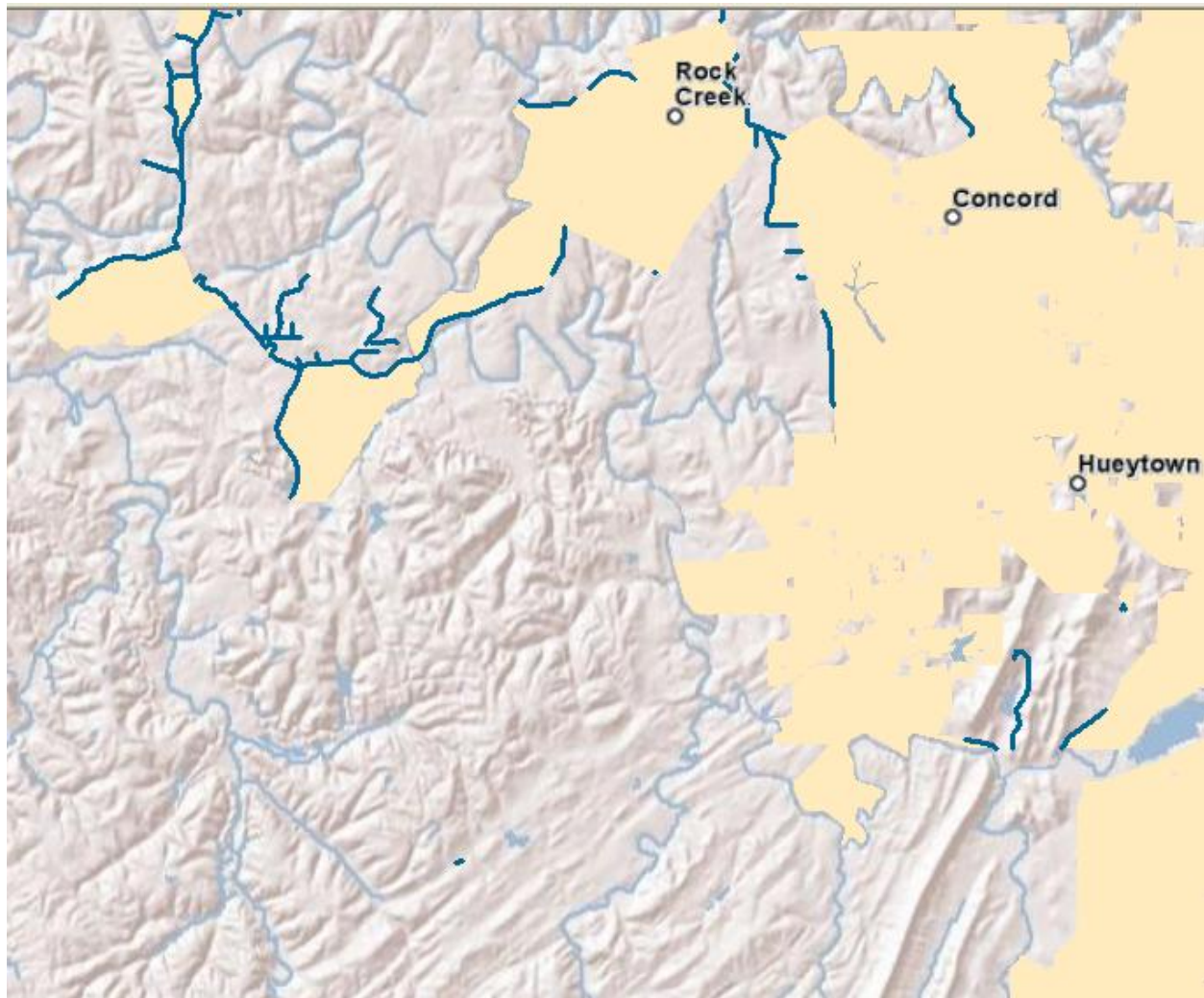


Figure 10-Blue road segments adjacent to peach covered small Blocks

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 edgemark our Block features and remain consistent with the Block size standards we used for other providers. 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.

<sup>13</sup> We continue to hear providers expressing concern that our request for either a geographic object or TIGER Line ID is beyond the scope of the NOFA clarification. Therefore, they cannot supply additional information to us.

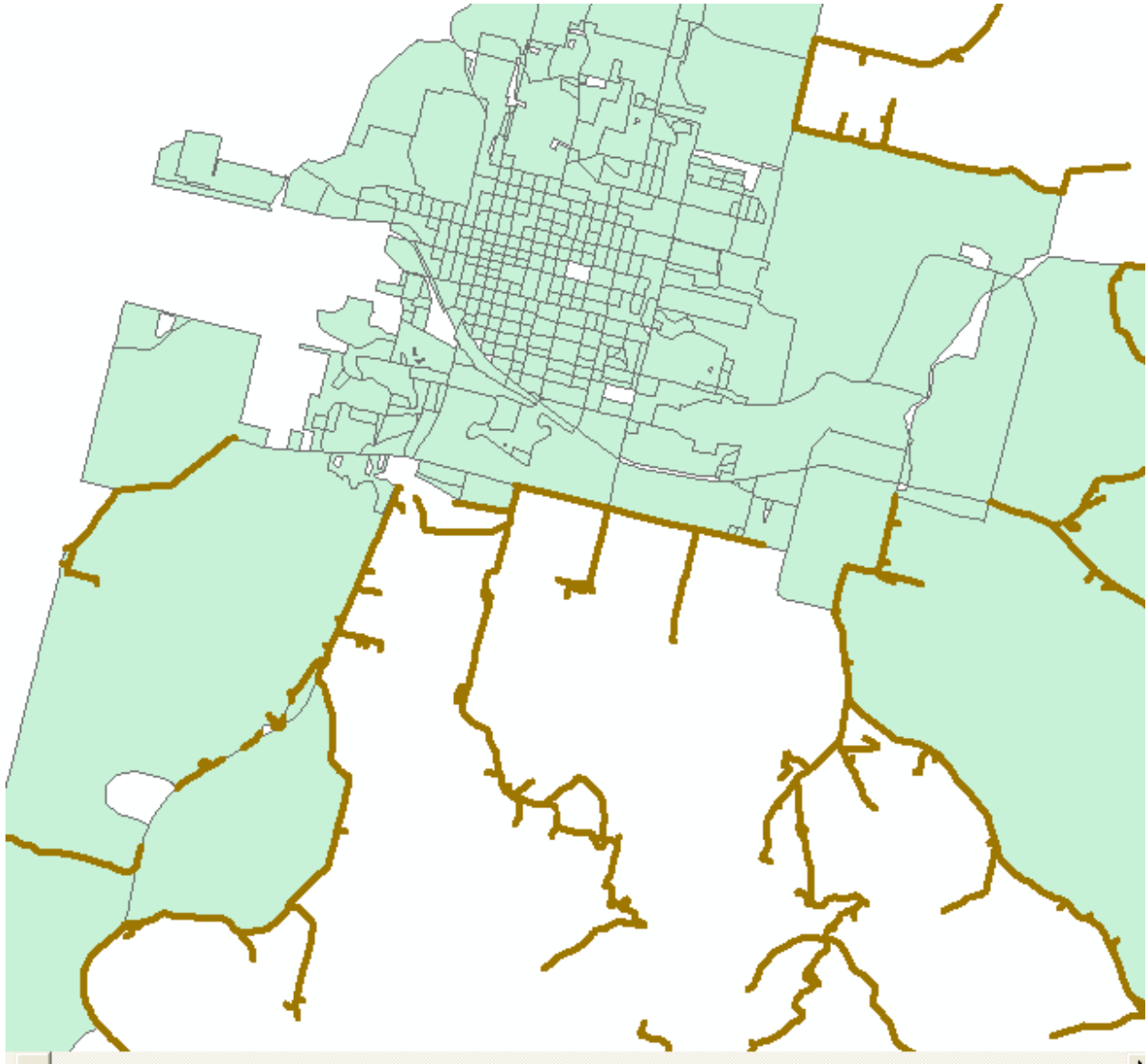


Figure 11-Provider Submitted Street Segment Objects. The segments don't edge match the Blocks nor are they continuous.

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

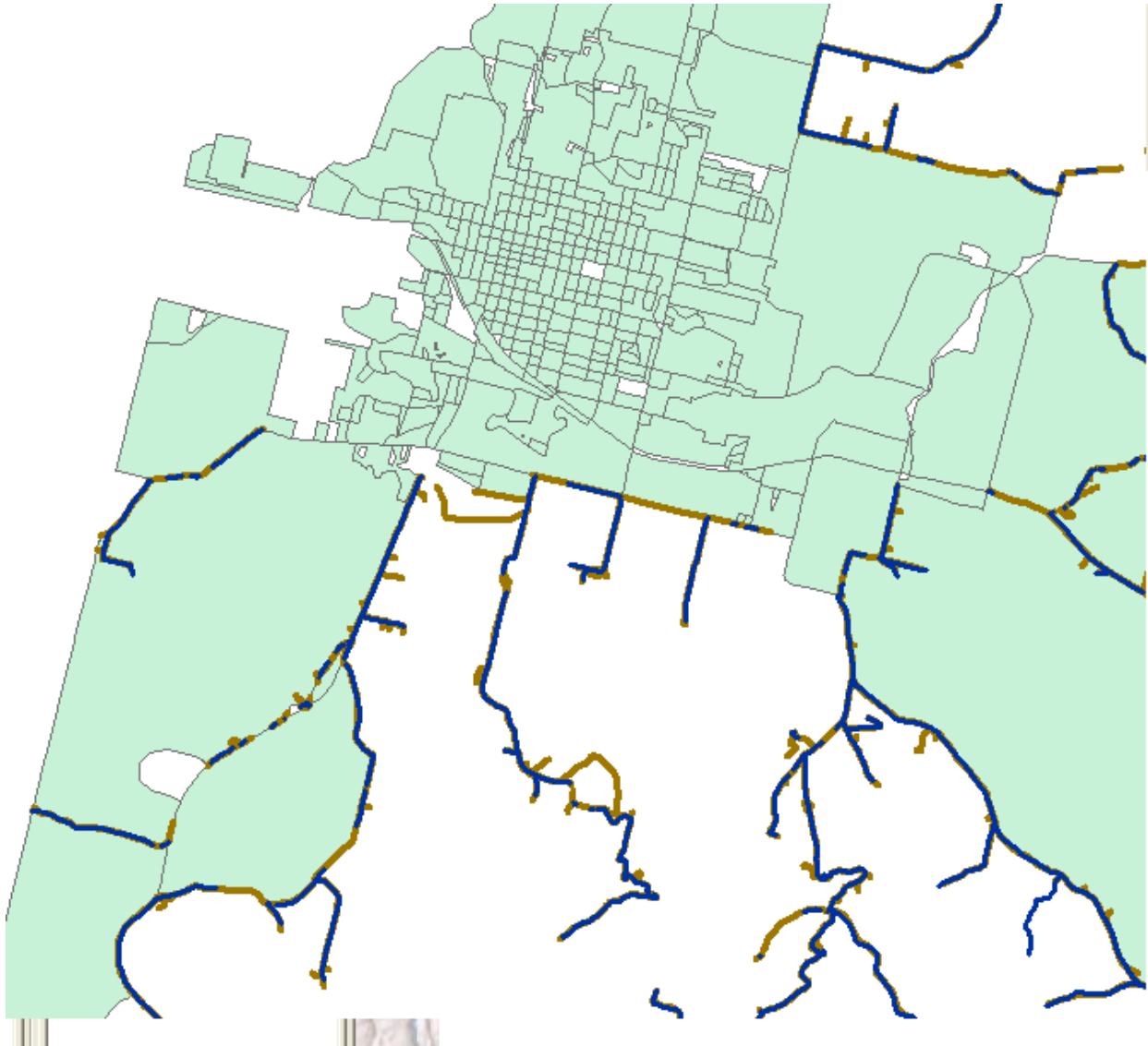


Figure 12-Provider submitted segments in gold, selected TIGER 2009 in blue—Conflation result; in many cases what was a continuous segment is made discontinuous because even with a distance buffer the TIGER segment doesn't always intersect the provider segment

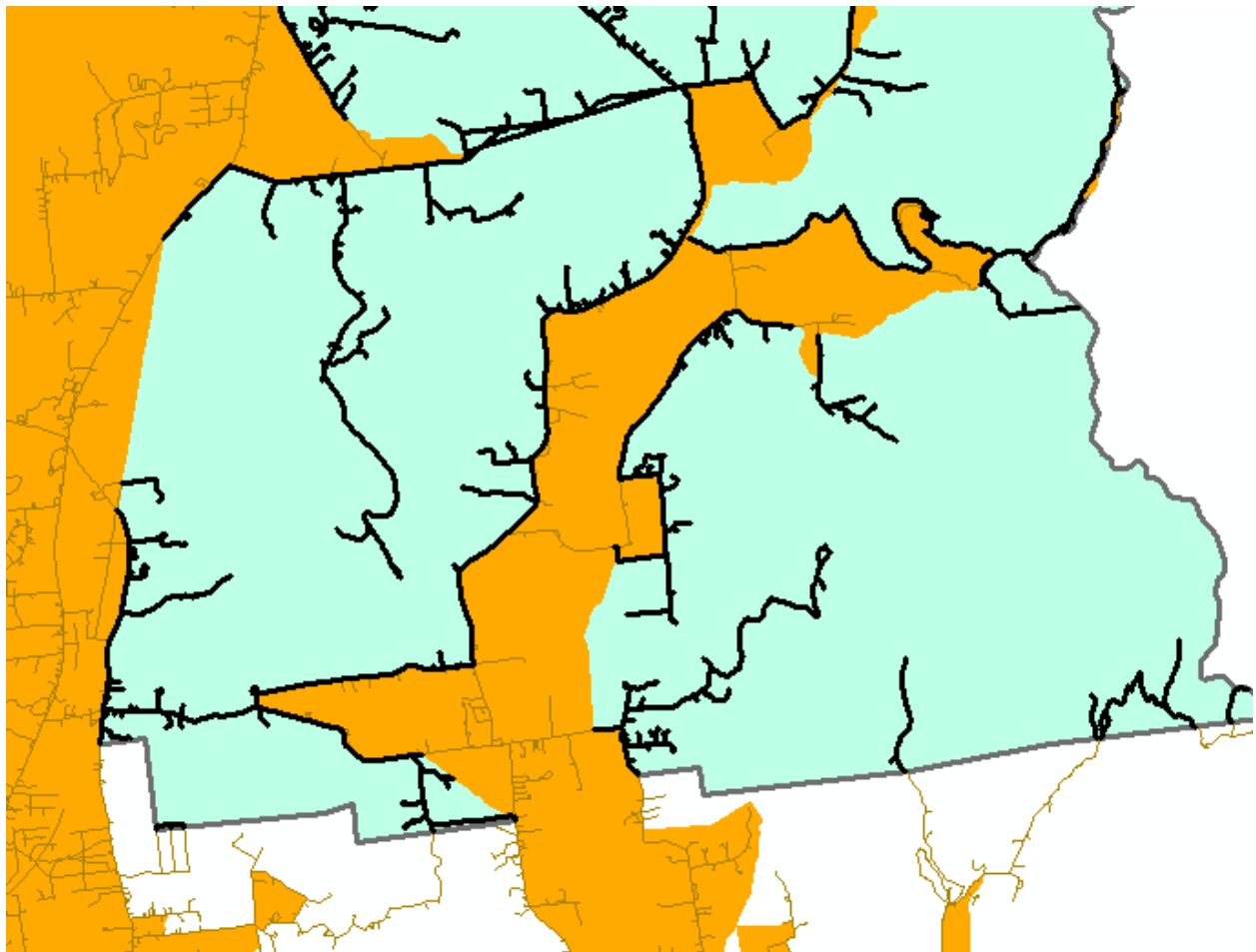
The final segment process was used when we were supplied with a Broadband covered area polygon. In this case, we found the segments within covered areas and eliminated those segments inside of Blocks less than or equal to 2.00 square miles.

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.<sup>14</sup>



**Figure 13-Output of the Segment Process**

### 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.<sup>15</sup>

<sup>14</sup> 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 these 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.



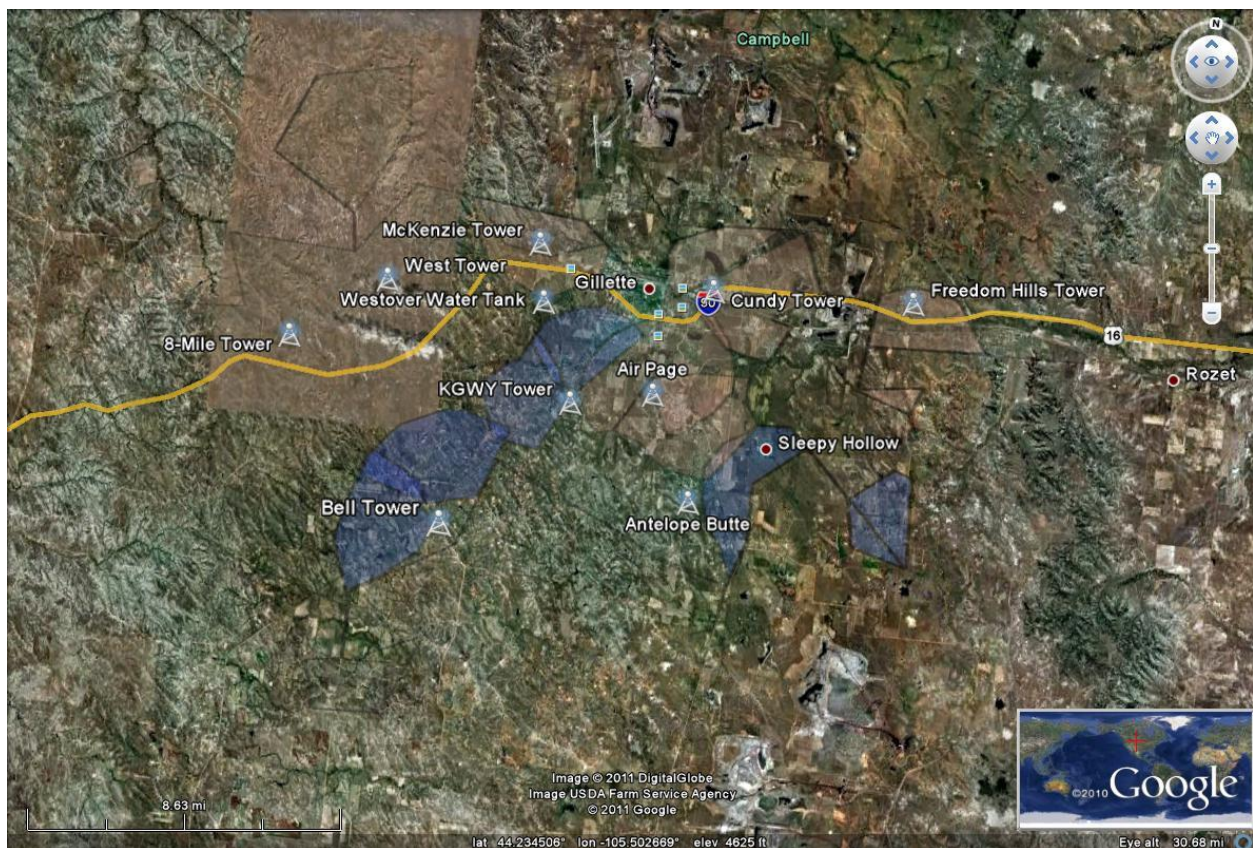
In this submission LinkAMERICA made an aggressive effort to bring additional WISP coverage into the NTIA dataset. For the most part, our outreach was with providers who were unable to supply sufficiently granular data in the past or those that could only submit wireless address points which is no longer a valid submission format.

In Round 3 fixed wireless providers generally either supplied coverage information or infrastructure from which coverage estimates could be derived. Many 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.

Other fixed providers were able to supply us with hand drawn maps or polygons/polylines drawn in Google Earth format. In these cases we did our best to georeference and verify the coverage areas with the WISP.

When we received coverage information in KML format, like the image below, we accepted the data as it was presented to us.



<sup>15</sup> Some polygon data did exceed the node count threshold. In these cases, data was rasterized to 100m cells and then converted back to polygons. The polygons were dissolved to multi-part geometry. This addressed the node count concern.

As the image above shows, in some cases we have hand-drawn coverage, as well as infrastructure. Instead of estimating their coverage using a line of sight or RF study, we elected to stick with the provider's supplied information. Our decision was guided by two primary factors:

- If the provider is advertising using this coverage they must have specific confidence in its accuracy.
- If the provider can supply coverage, as well as infrastructure that reasonably supports the coverage, there is a very high likelihood in the accuracy of the information.

The downside, of course, is the polygon shown on the map may not represent our notion of how wireless coverage should appear.

In general we note several interesting trends in the wireless data. First, we can be successful in increasing the amount of WISP coverage when we aggressively pursue WISPs. This means we have to be willing to accept data on their terms and convey it into SBDD formats. Some of our WISP submissions have taken over 12 hours to normalize into SBDD formats. Second, we have to accept that some WISPs will not be able to supply FRNs. There remains a minority of WISP providers who are not aware of the FCC FRN. Third, 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. Fourth, it was very difficult getting providers to identify spectra used for Broadband data services<sup>16</sup>. 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 one mile of a boundary are not submitted to NTIA.

We retain from the provider the provided latitude and longitude, as well as Census block. For some coverage data, if a provider is unable to supply a longitude, latitude or Census block, we fill in these attributes. In those circumstances where we do not have a Census block, but we do have a longitude

---

<sup>16</sup> One provider responded by email, "This mapping program is to provide the coverage area for Broadband provided by a company. Not to keep a detailed account of every aspect of a companies (sic) network."

and latitude, we accept the given longitude and latitude and use that as the basis for our Census block assignment.

With point data we have tested for comparable geocoding success rates but do not overwrite provider information. From this type of analysis we note the amount (usually little 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.

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<sup>17</sup> 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<sup>18</sup>. 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 scrape 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<sup>19</sup>. Because much wireless propagation is driven far below the Census Block and much engineering information isn't

---

<sup>17</sup> 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.

<sup>18</sup> Central Office location was derived from MapInfo ExchangeInfo Professional. Wirecenter boundaries also came from this commercial product.

<sup>19</sup> 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.

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.

### 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 to populate typical speed columns.

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 value.

Several 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.

We continue to request advertised speed at the block level. Nevertheless we appear to be getting speeds that do not vary over a large geographic area – leading us to believe that providers may still be submitting the maximum speed advertised in local media for the entire market. For the most part, we have been unsuccessful in messaging 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.’<sup>20</sup>

In circumstances where a provider supplies a range of speed attributes, we assign NTIA categories based upon the midpoint of the range.

To support NTIA program office requests, we have also modified the structure of the Service Overview table. Even if Maximum Advertised Speed is supplied at the market or county level, we push that speed down to the contained Blocks. The only records that remain in this table, will be those wireline records with either a non NULL nominal weighted speed or ARPU value.

## 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 the second submission we

---

<sup>20</sup> 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.



continued to obtain additional connectivity information. For the Spring 2011 collection, the team began a survey process to directly engage these important organizations.

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 has strong ties into the 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 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 possible. 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 that stand to produce the greatest synergies with the SBDD planning process.

The above 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. In another state we are attempting to glean connectivity information from a municipal government survey. There may be some downside to this collaborative approach in that we may have to work with data spanning different times or we may not have all of the location-specific information we need, but this does prevent the same user from receiving multiple inquiries.

Further, the team continues to rely on the notion of Internet Intensity Zones. As the Broadband coverage information is developed, if we do not have definitive connectivity information from other sources (e.g. a phone survey, web survey, listing provided by a facility owner) in this study, those Anchor points that fall into an existing area of SBDD Broadband coverage will not be left out or submitted with NULL values. 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.<sup>21</sup>

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

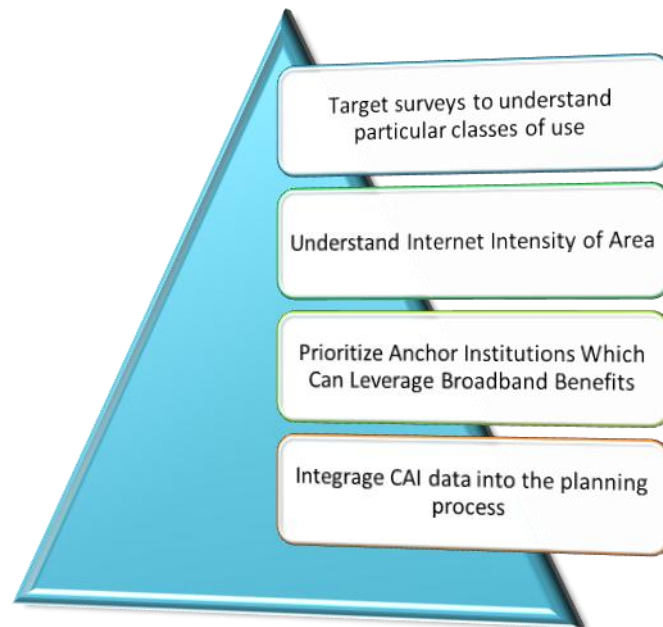


Figure 14-Anchor Institution Process

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<sup>22</sup>, this likely means that the Broadband accessibility is very comparable for CAIs that are close together. We believe this means Broadband accessibility may be less about connectivity than it is about the ability of a CAI to afford, 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.

<sup>21</sup> 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.

<sup>22</sup> [http://en.wikipedia.org/wiki/Tobler's\\_first\\_law\\_of\\_geography](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..

## Anchor Institution Survey

During the third submission period we began a survey process to both verify received connectivity information and garner additional connectivity information from CAIs. As with WISPS we wanted to aggressively target and improve this data section.

The process began with the Round 2 CAI list. Again, we prioritized schools, libraries and healthcare institutions. A small team made outgoing phone calls to discover relevant contact names. In Wisconsin, we were able to gather about 150 email addresses based upon 440 calls. There were only 14 refusals.

While one team worked on improving the contact list, a second team designed and developed a simple online survey system called CAVS (Community Anchor Verification Survey).

---

**Anchor Name**  
CAVS TEST CAI

---

Please use the fields below to enter your organization's address. We are interested in the physical location of your organization. If you have both a Post Office box as well as a street address, we would prefer the street address.

<b>Building Number</b> <input type="text"/>	<b>Street Prefix</b> <input type="text"/>	<b>Street Name</b> <input type="text"/>	<b>Street Type</b> <input type="text"/>	<b>Street Suffix</b> <input type="text"/>
<b>City</b> <input type="text"/>	<b>State</b> WI	<b>ZIP 5</b> <input type="text"/>	<b>ZIP 4</b> <input type="text"/>	

---

**Category**  
Medical/healthcare

**Does your organization currently subscribe to broadband service?**  
No

**How does your organization receive broadband Internet access?** [Broadband Technology Descriptions](#)  
All Other

**What is the maximum available upload speed of the Internet connection at this location?** [Conversion Table](#)  
Greater than 200 kbps and less than 768 kbps

**What is the maximum available download speed of the Internet connection at this location?** [Conversion Table](#)  
Greater than 200 kbps and less than 768 kbps

Figure 15--CAVS Screen

Users were invited into the CAVS system by the receipt of a postcard with an organization specific code printed on the mailing label. Beyond the questions shown above, there was a second page to the survey dealing with use of Broadband. Those results are directed to the planning teams.

The table below summarizes outgoing contact activities by state. This includes both a post card as well as for some organizations in which we had contact information a follow up phone call.

States	Post Card	Calls
WI*	2033	75
ID	1059	259
WY	345	30
AL	1640	14

As of 3/16, verification<sup>23</sup> statistics were as follows:

State	Verified / Total Records	Percent Verified
AL	72/2137	3.3%
ID	172/1596	10%
WI <sup>24</sup>	1187/3945	30%
WY	169/796	21%

We are keeping the survey open after the Round 3 submission to NTIA and will continue to collect data. In Alabama we have also begun to use resources from the planning teams to make outgoing calls and better target the surveys.

Clearly this survey was resource intensive but it did yield an increase in verified, rather than estimated, CAI data. We are unsure if we can sustain it in the next submission, but it has proven to yield new information.

### Anchor Institution Trends

At this point we have focused our CAI attention on schools and libraries, 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.

<sup>23</sup> We say a record is verified when it has been opened by the CAVS test user. It means at least one field was modified.

<sup>24</sup> In Wisconsin several large school districts supplied files with connectivity information; we performed a bulk update in these cases. We attribute it to the survey as the survey triggered this response.



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.

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.<sup>25</sup> At this point we have had minimal success gaining this information.

Because we have a wide ranging population of CAIs in our data set we have a 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.

As a final verification step, we attempt to screen the CAI data for duplicate values. Because many CAI are closely clustered together we perform the de-duplication based upon the ANCHORNAME within the ZIP code.

## 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.

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.")<sup>26</sup>

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 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).

---

<sup>25</sup> Within the public safety category, it is also very difficult to derive precise locations as many CAI are addressed to PO boxes.

<sup>26</sup> From [http://broadbandusa.gov/files/BroadbandMappingNOFA\(FederalRegisterVersion\).pdf](http://broadbandusa.gov/files/BroadbandMappingNOFA(FederalRegisterVersion).pdf) at 54, visited March 28, 2010

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<sup>27</sup>. Others would not disclose as their Middle Mile points are not 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.

Additionally, based upon the new Provider\_Type classification of "other," we have started to integrate points provided by Broadband service providers not meeting the NOFA definition. This includes POP locations and aggregation points for public / private networks.<sup>28</sup> Within a given submission there were two final attributes that tended to concern respondents. First, speed should 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 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 a sample of 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 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. To a limited extent we also use licensing locations and tower infrastructure to spot-check supplied

---

<sup>27</sup> 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}"

<sup>28</sup> As discussed in our readme.txt file, a number of middle mile points were lost in validation due to their location in adjacent state. This will cause a decrease in some providers relative to prior submission.

coverage. This mode of verification remains complex, given the lack of facility-based information with mobile wireless.

## 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 of our data verification routines. Consistent with the movement toward an fGDB export database and use of a data receipt script, much of our validation effort was spent in supporting the ETL processes into the required formats. In future data submissions we will continue our work to stabilize and improve the business process that normalizes provider submissions into NOFA formats and expands 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.

We are learning that Verification has multiple dimensions.

Provider verification is finding providers who supply Broadband and discriminate out providers not meeting Technical Appendix A's definition of Broadband.

Identity verification is taking the provider's categorized in the first step and ensuring that the provider either has a valid FRN or is assigned a default FRN. Identity verification is very complicated because of the Technical Appendix A's mandate to record data at the FRN, Provider Name and DBA level. Each of these attributes could be unique for a single provider going to market under different or the same names. As a result, rolling up each provider into an identity collection that matches either the FCC data integration team or a third party Broadband provider's data view, is very, very time intensive. Identity verification is discussed in the earlier section-- Developing the Provider List.

Coverage verification is a broad term, but in our definition it boils down to determining if Broadband coverage is in the right place. For a given provider, the question is whether the coverage is assigned to appropriate Census Blocks, road segments or area features. Coverage verification can be further broken out into two distinct classes:

- Technology verification, which is determining if the provider is listed with a technology consistent with their marketing information. It also involves a validation with supplied speeds.
- Speed verification, which is determining if the speed supplied for that block, road segment, point area file or market area is consistent with the technology and the marketing information received.

The final verification dimension is consumer feedback and crowd-source verification. This is a dynamic set of steps we are beginning to implement. One side of this is responding to consumer concerns. The

second is using the crowd sourced data to validate provider claims and, if appropriate, update the map and the underlying data.

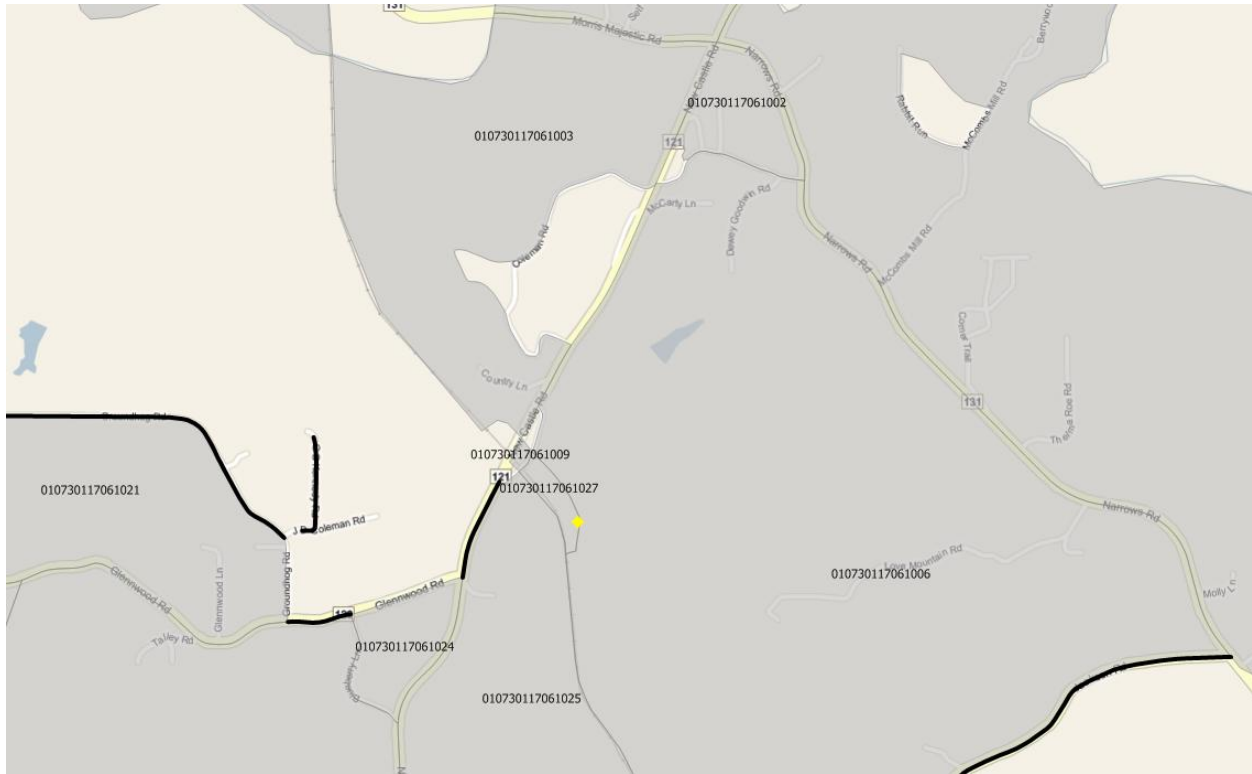
At this stage, our working hypothesis (confirmed by our experience) is that there will not be a single dispositive measure to indicate Broadband coverage availability in a Census block or along a segment. From prior work, and examining our current provider submissions, we believe that there is too much 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 both relevant to and 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, coverage verification is at the level of data submitted to NTIA.
- Second, coverage 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 coverage verification process to implement is the erroneous dispersion of coverage. These are the “islands” of coverage isolated by significant distance from other covered areas. This is the opposite of the Internet Intensity Zone notion discussed in the Community Anchor Institution section. In other words, Broadband Internet likely doesn’t exist far away from other areas with Broadband Internet access.

Before explaining our overall verification thought process, we have several examples, which illustrate the complexity of coverage verification.

The first example 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 16--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.

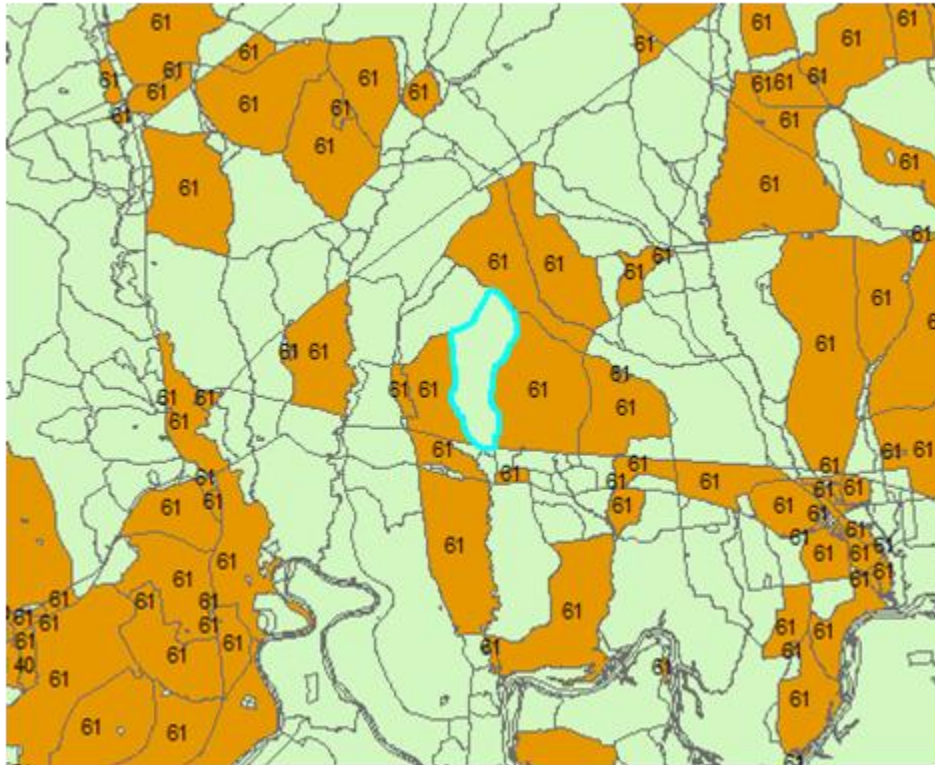


Figure 17--Dispersion in Submitted Data

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.



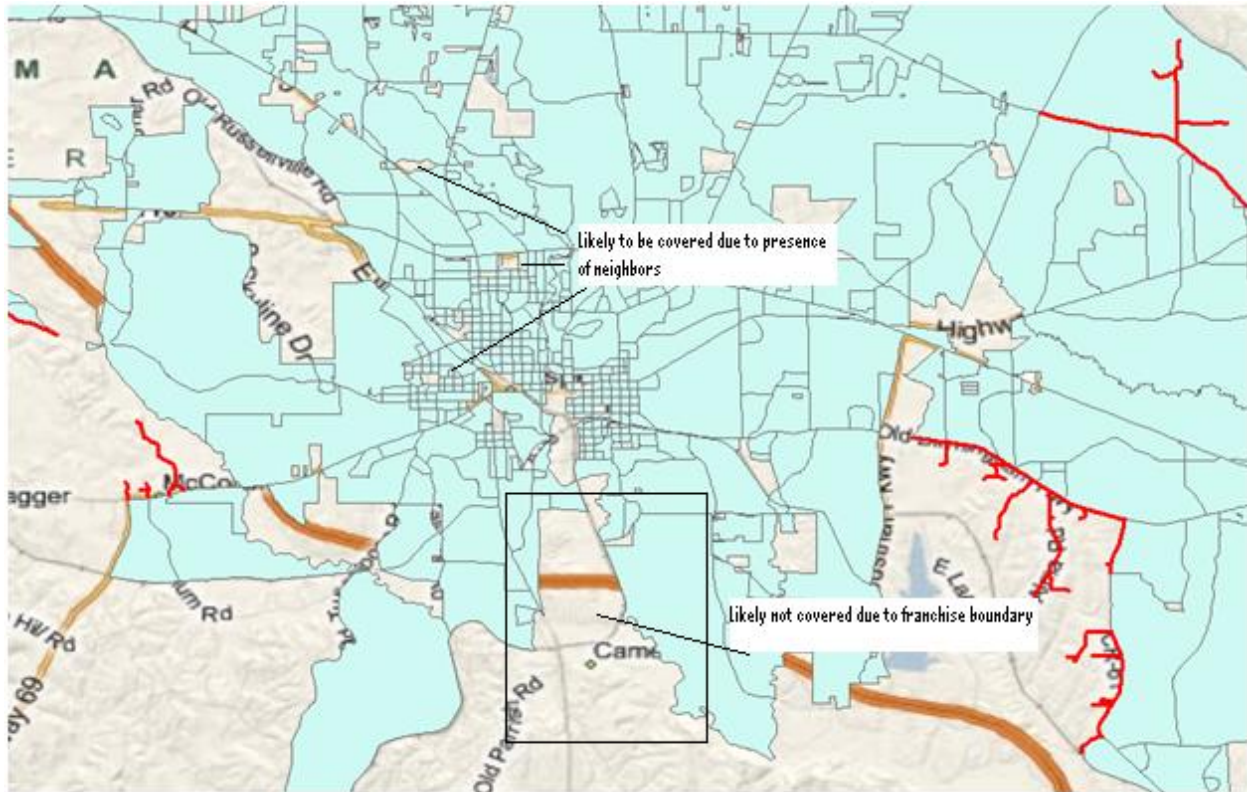


Figure 18-Where do you stop interpolating?

From what we can gather from some providers, the submitted data—data with consistently high degrees of dispersion or coverage holes—tends to come 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 contend with this type of circumstance. However, we have not been entirely successful at getting franchise boundaries that would address much of the issue.

The final map shows this dispersion problem, but to 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 NOFA clarification.



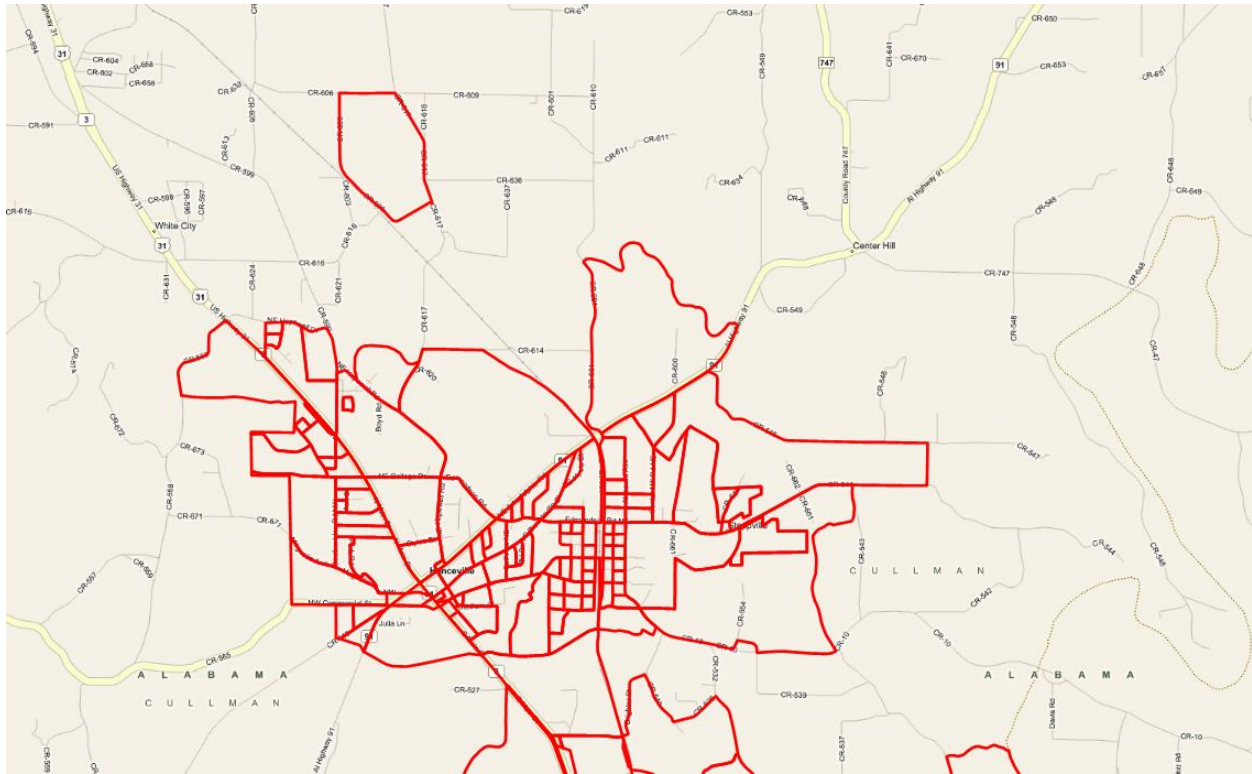


Figure 19-Dispersion in covered Blocks

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 DSL providers because we have a good source of exchange boundaries.

The exchange boundary dispersion verification method breaks down when examining smaller providers who are more likely to CLEC into neighboring territory. In the figure below, the black line represents the exchange boundary, while the continuity in the DSLAMs likely points to coverage extending along a road into another provider's territory.



Figure 20--DSL Coverage outside of exchange boundary

In sum, the variability in our source data continues to suggest that our dynamic verification process is relevant, appropriate and evolving in a manner consistent with the overall program. And, as noted above, we believe the more meaningful outcome of our verification processes 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 given the variation in our input data and the varied perceptions of service providers, map viewers and down-stream data consumers.

### **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 any immediate questions or concerns are sent back to the provider as quickly as possible. 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 second 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 April, 2011 Round 3 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 dispersed 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 submissions have moved online, we have also begun to benefit from crowd source 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 coverage requirement, we continue to receive complaints that the coverage shown on the map is not accurate for a particular location within that block.

### **Consumer and Provider Responses to Deliverables**

Here, we segue from internal verification to external verification. We view responses to our work product as a form of validation and verification. On the one hand, this gives us the opportunity to fix mistakes and then generate QA steps to make sure that the problem does not reoccur. We also learn how to improve what we are doing or better explain what we are doing to a community not always familiar with the NOFA and program office framework. On the other hand, listening and learning from

this feedback helps us better target our mapping deliverable to meet the needs of our external customers. In this second case, external feedback not only provides feedback on perceived qualities (or lack of quality) in the data, it helps us to learn if we are developing data that is truly helpful to downstream users.

At this point, our external deliverables take three forms: State Broadband Maps, data transfer to NTIA used for the National Broadband Map, and text format data requested by outside parties.

### **Online Map Experiences**

Now that our State 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 strong concerns regarding how wireline services are displayed, as contrasted to how wireless coverage is displayed. This is an artifact of the SBDD data model. As an example, consider the figure below.

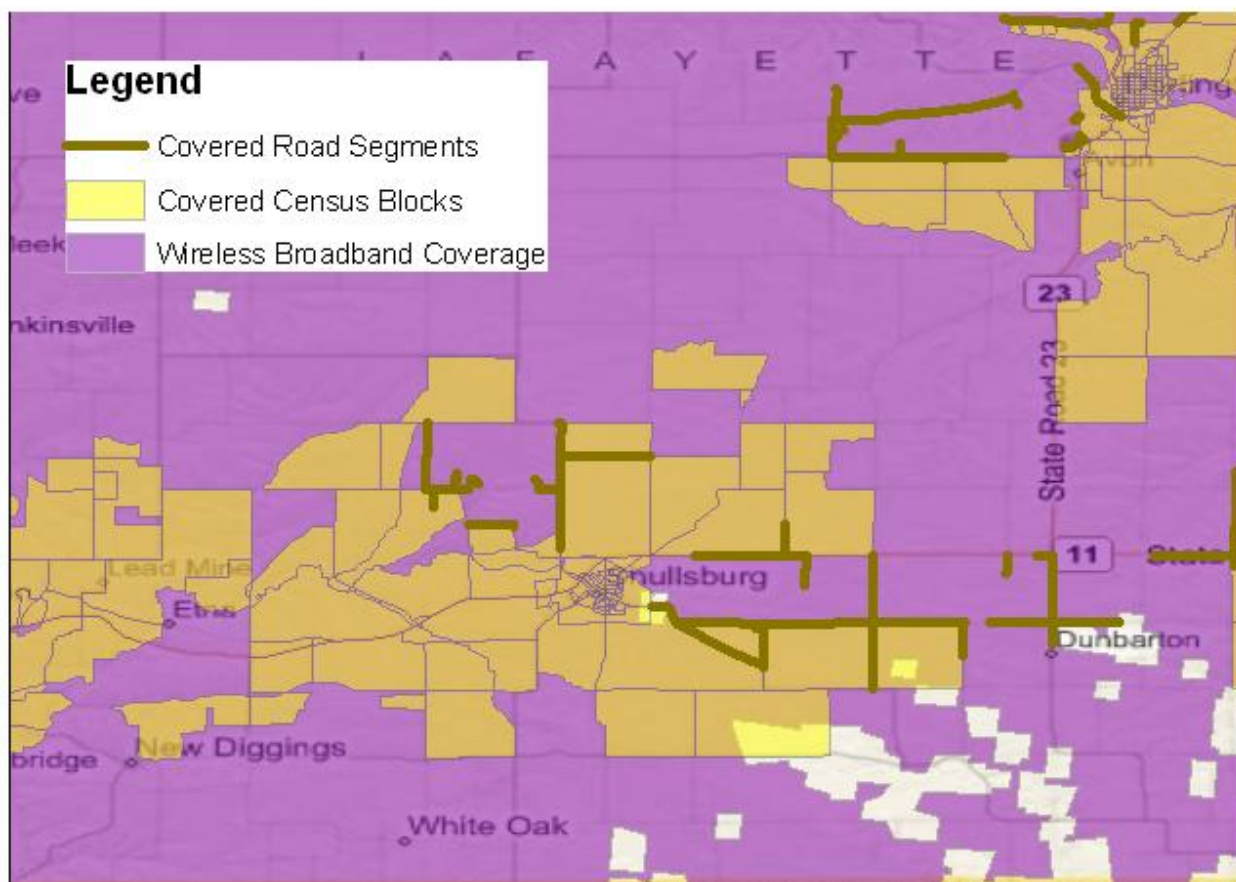


Figure 21--Multi Network Coverage portrayal

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 because 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 because 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 have expressed 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. Their ability (or lack thereof) to actually provision such services for new users within a 7-10 day period adds yet another level of complexity when attempting to fairly portray their coverage capabilities.

### 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? In the past, we reported back to the Census block, or buffered road segment intersected by the user click. As far as the map was concerned, once we move off of that road, or out of that segment, we have a new area to examine.

Our sense, given feedback received, 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 general 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. As of this date we cannot determine if the claims of inaccurate coverage in online maps are due to the looser provider view standard or not. We keep this looser standard in place to minimize the likelihood of self-disqualifications.

### National Broadband Map Experiences

When the National Broadband Map launched, our phones began to ring.

Responding to a number of provider inquiries as well as emails from citizens provided some insights. It also illustrated that we now bear a second dimension of external verification. That is, we must be prepared to respond to people who are confused by apparent inconsistencies between the State and National Broadband Maps<sup>29</sup>.

The case below, based upon a call we received, illustrates some interesting intersections between the State and NBM.

In this example a Citizen called inquiring about the difference in results between the National Broadband Map and our State of Alabama map. The issue in question was coverage at his home. The Alabama map showed he had coverage at his home, but the National Broadband Map said he did not.

In the image below, the green dot represents the geocoded location of his home. Based upon imagery, the geocode is quite accurate. The olive colored polygon represents a covered Census block less than or equal to 2.0 square miles. The Census block shows coverage by a number of wireline providers.

The geocoded point is about 170' from this covered Census block.

---

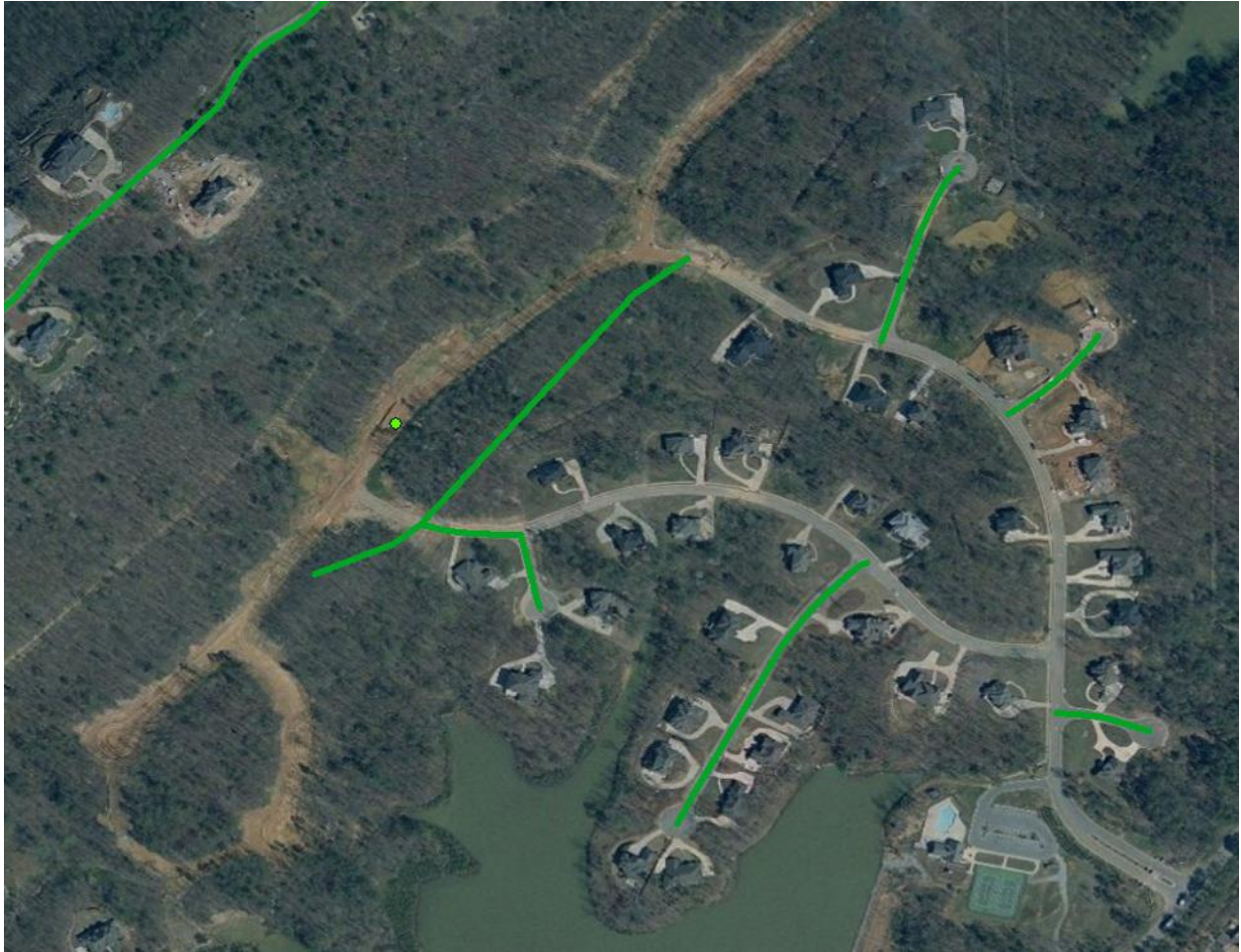
<sup>29</sup> We have a similar concern regarding textual data extracts. We may translate our SBDD submission into covered Census Blocks in a way that is different from NTIA.





Figure 22-NBM Covered Census block example

In the next image, covered TIGER road segments are shown in green. It is important to note how far the TIGER road centerlines are from the actual roads in the subdivision. It appears the geocoded point is reflecting more recent and more accurate road centerlines, placing the green dot at the correct location. Since the SBDD data is submitted in terms of TIGER 2000 the road on our map shows up about 100-200 ft away from where that road is located today.



As mentioned previously, however, our online maps buffer road segments to 300 feet on either side of the road centerline. In this case then, our state map buffer is large enough to return valid service providers for this green dot. The NBM, on the other hand, does not appear to buffer segments or the edges of census Blocks and will not return providers for this location. Our intent in this example is not to criticize the national map; rather, it is to illustrate that we may inadvertently make trade-offs between false positives and false negatives, differently.

This case illustrates several important tensions between the data as we present it to NTIA, map it ourselves and because of how it may be viewed within NBM context. A lack of agreement on how to handle these inconsistencies in the source data and differences in mapping approaches may cause consumer confusion.

The issues seem to come down to this

- a) How do you (or can you) handle the impact of time when roads move between TIGER versions or between TIGER and other road products? In this case, online map road traces will not show up in the right area.



b) Given the inconsistencies between TIGER geometry used in submission and underlying roadbases used for geocoding online, how do you (or should you) insulate the viewer from the inconsistencies. There appears to be a strong likelihood that TIGER judges a particular point to be in a larger than 2.00 sq mile Census block while that same location could be in a small block area in the online view.

c) How much tolerance should be introduced when returning a list of valid providers? Is it better to error on gathering too many providers or too few?

d) Since the NBM gathers feedback based upon its representation of coverage, how can/how should this crowd sourced feedback influence data presented in a different manner elsewhere?

## 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 2009 or alternative (TeleAtlas, possibly) Blocks. Especially with the need to derive segment geographies, we would prefer to message the 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 Fall 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 have asked CostQuest/LinkAMERICA not to 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 a Broadband Provider is—and how that definition 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<sup>30</sup>. Further, the need for clarification around a facilities-based provider, versus the reseller, has injected even more ambiguity into the mix. Right now we are unclear on how

---

<sup>30</sup> 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 turnup 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.

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.

Again, we do not want to exclude a service provider, but we believe there needs to be further clarification around the 7-10 day "rule," the definition of a "reseller," and better interpretation of facility-based providers, versus equipping UNEs, SpA or leased lines.

We have used the Provider Type of "Other" to classify a number of providers who offer Broadband services, but we do not offer them in a manner consistent with Technical Appendix A definitions.

### **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 calling 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<sup>31</sup>. 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 error on overstating coverage to 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 facility-based evidence from which to remove it?

### **Granular Data Collection Issues**

#### **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

---

<sup>31</sup> 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. In a matter of speaking, the Provider\_Type is more determined by the technology and/or location than a type of business. The core intent of the NOFA and our grant application was centered around the 7-10 day providers but we believe maintaining information on Provider Type "Other" and "Reseller" is important to assist in validation and market segment analysis as resources are available.

was used to georeference plant (is it the carrier's roadbase?). This leads to uncertainty regarding the input data scale or accuracy of 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 December 31, 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, then the data created under the NOFA will not 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.

### 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, while 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 of how the information will be used, or 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 will then 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.). Finally we are also noting a significant increase in speed reported to us. This may be due to network upgrades or competitive concerns to match the theoretical network speed.

### **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 as possible, to know what Census Block vintage 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. As a departure from past practice, where a provider submitted Middle Mile points out of state, we are no longer passing those points to NTIA as they fail the validation script. We experienced validation errors for BroadbandServed=N records in the CAI table. These records were attributed a Technology of Transfer=0. This cleared validation.

5. In tables with mandatory Zip5 (Service Address), if the End\_User\_Zipcode was not available, we have inserted '00000'
6. We have a significant amount of VDSL, ADSL 2 and ADSL 2+ coverage categorized into the xADSL category.
7. 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.
8. 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).
9. We note a few providers who have speeds seemingly inconsistent with their technology of transmission. This is either very low speeds with 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.

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



Access constraints: <a href="#">None</a>					
Use constraints:					
There are no restrictions on distribution of the data by users.					
<b>Census Block</b>	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
Access constraints: <a href="#">None</a>					
Use constraints:					
There are no restrictions on distribution of the data by users.					
<b>Service Overview</b>	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: <a href="#">None</a>				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
<b>Road Segment</b>	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
	Access constraints: <a href="#">None.</a>				
	Use constraints:				
	There are no restrictions on distribution of the data by users.				
<b>Wireless</b>	Constraints on accessing and using the data	No	Yes	Yes	NO attributes on any record in this feature class are considered confidential
	Access constraints: <a href="#">None</a>				
	Use constraints:				

There are no restrictions on distribution of  
the data by users