State Broadband Initiative Mapping Methodology

For the States of Alabama, Idaho, Wisconsin and Wyoming
Revised September 30, 2011

CostQuest Associates

LinkAMERICA Alliance

Copyright 2011, All Rights Reserved

Contents

Overview

The following documentation provides an overview of how the fourth required data set was collected and processed for the State Broadband Initiative (SBI) 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 changes based upon NTIA guidance
    a. Modification of Satellite providers as a Type 1 Broadband provider;
    b. Discontinuation of estimating Community Anchor Institution coverage and speed;
    c. Review of submitted speed with respect to NTIA supplied frequency table
4. Transition planning with respect to capacity building within the State for Broadband map development
5. Development and posting of a provider Type classification rubric

Treatment of the following subjects has been expanded:

1. Community anchor institutions and survey methodology
2. Verification and validation
3. Data production methods
4. Conversion to Census 2010

As anticipated, the SBI program continues to mature and evolve. Technical leadership and strong program office 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 April 1, 2011 “Round 3” 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-subsidiary structure of many telecommunications companies, make provider identification and tracking very challenging.

In early July 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 4", this effort continued on a daily basis until we reached our final data submission deadline on August 19, 2011. After August 19, 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 4 providers that responded too late to be included in the final dataset will be included in our Round 5 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.[1] 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 Four. 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 previous rounds they were offered the opportunity to do so in this collection round. 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 all cases, minimum standards established by the NOFA are honored. In other cases, providers chose to submit data without executing an NDA.

Provider Survey

Since three 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 4. This allowed us to provide two response options to providers. The first was for them to review check maps of their coverage and speed data – submitting only corrections and additions to the existing dataset. (For provider convenience the

[1] As with other Grantees, we struggle with appropriate and consistent classification for service providers who opportunistically provision Broadband services. In this submission we continue 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.

check maps were created in both PDF and Google Earth (.KMZ) formats.) 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.

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 4 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 previous rounds, 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: For Round 4 we continued to work with participating providers to improve their datasets. The change in Census Data vintage was explained to providers and links to appropriate files were provided to assist with the transition to the new vintage data.

Check maps continue to be a useful tool to show providers how their area would be displayed on the resulting interactive state map and to get constructive feedback regarding corrections and changes that need to be made to their coverage and speed data. 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 4 survey in early July 2011, LinkAMERICA launched an extensive effort to encourage responses. Every known provider was contacted at least twice during the months of July and August. The initial data submission deadline was set for August 19, but, as previously noted, we continued to accept "straggler" submissions into September.

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 providers who have simply ceased to exist[2].

[2]The 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 4, given the data vintage of June 30, 2011, we began this process in July and the last submissions were accepted in September, 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 SBI 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.
- FCC 477 restricted use data were analyzed to find valid providers within a given area.

We have included third party data sources, which touch on each of the three major technologies analyzed within the SBI 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.[3]

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

a. Geographic Data Files

[3] We also suggested forming a technical standards committee and a consistent system for confidence reporting.

i. US Census TIGER data[4]

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 Coals reports (321/325)
   v. FCC FRN API lookup tool
   vi. FCC/FAA Antenna Registration System
   vii. FCC FRN Lookup Tool (plain text search)
   viii. 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
   vi. List of museums, conventions, and visitors bureaus from 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, when submitted data conflict with prior submissions 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.

[4] Census data were derived from < http://www.census.gov/cgi-bin/geo/shapefiles2010/main>, Census 2010 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.

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 time tested telephone survey approach focused on the consumer market. In addition, we continue to advance our process to include certain initiatives centered in two social media outlets – Facebook and Twitter. These initiatives are discussed below.

Consumer Surveys

Working under contract for the state of Alabama in 2009, our initial consumer survey was performed before the NTIA SBI grant was in place. Subsequent consumer surveys funded by the SBI grant were hosted in 2010 for the states of Idaho, Wisconsin and Wyoming and then again in 2011 for Alabama (as noted below). These surveys will be repeated after two years to establish and evaluate trends. To this end, in August/September 2011 we are wrapping up a second-round survey in Alabama designed to expand our understanding of important adoption issues and to establish important local trends from the initial 2009 survey. Survey results from this effort are currently under evaluation. 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. In the most recent Alabama survey we added a third tier to our approach as we equipped local field survey teams with an iPad-based survey tool and targeted their time to reaching the younger market. 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 updated consumer survey in Alabama was designed to produce results relevant to the state's twelve Broadband planning regions).

The resulting data is helpful on a number of fronts in the SBI'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 SBI 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).

Our ongoing consumer survey process adheres to a consistent process. For example, consistent with prior practice the 2011 Alabama survey was launched in June 2011 with a test number of survey calls to confirm (and adjust as needed) the structure of the survey and the underlying survey process. Our surveys typically run for three to four months. All telephone surveys are completely random beginning with the acquisition of a list of state-specific, randomly selected landline telephone numbers. Mobile phones are not typically included in the surveys. Upon evaluation of the survey statistics, auxiliary surveys are executed to ensure appropriate representation is achieved on both demographic and geographic fronts. For example and as noted above, the recent Alabama survey was augmented with a field effort to ensure the younger demographic (i.e., age 18 – 25) was adequately represented. This secondary step is required because of the continued migration (by younger markets) to non-landline based communications. This younger market is also surveyed by reaching out through social media outlets to encourage their participation in an online survey process.

Survey statistics from the Alabama update survey are currently being developed and evaluated. Survey statistics from our initial surveys in Idaho, Wisconsin and Wyoming were summarized in our last filing. Survey volumes are designed to achieve statistical validity.

As noted above, our telephone survey process is augmented by providing online access to the survey. Participation in the online survey is 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, these tend to be long surveys. The surveys typically average just over fifteen minutes. While this clearly contributes to the number of survey call attempts that were required to reach the level of statistical validity, it is 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 SBI efforts. All of our efforts are informed by and consistent with relevant state statues 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 LinkedIn, Facebook and Twitter to support our work. A central focus is on promoting awareness. As noted above, we are able to promote additional input on the consumer surveys through a social media outreach program aimed at our younger market segments.

In addition, we continue to evaluate 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.

Capacity Building and Transitioning to State Partners

A foundational goal of LinkAMERICA has always been to transfer knowledge and capacity to our State partners. As we move into program year 3, distinct tasks are migrating to the responsibility of our State partners.

Within each State, transition planning and responsibility for specific activities is on a slightly different timeline. Much of this is driven by resource availability and partner identification within the State. For example in round 3, the State of Alabama used interns to validate Community Anchor Institution (CAI) data. In this submission Alabama took on greater responsibility for the CAI submission. To support this LinkAMERICA developed a detailed transition document describing the current CAI efforts.

Other States are looking more towards program year 3 and the in-State hire of a Broadband Coordinator as the initiation point to support their transition efforts.

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 to take advantage of increasing specialization and proficiency with certain types of data and management responsibilities. In submission 3, 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.

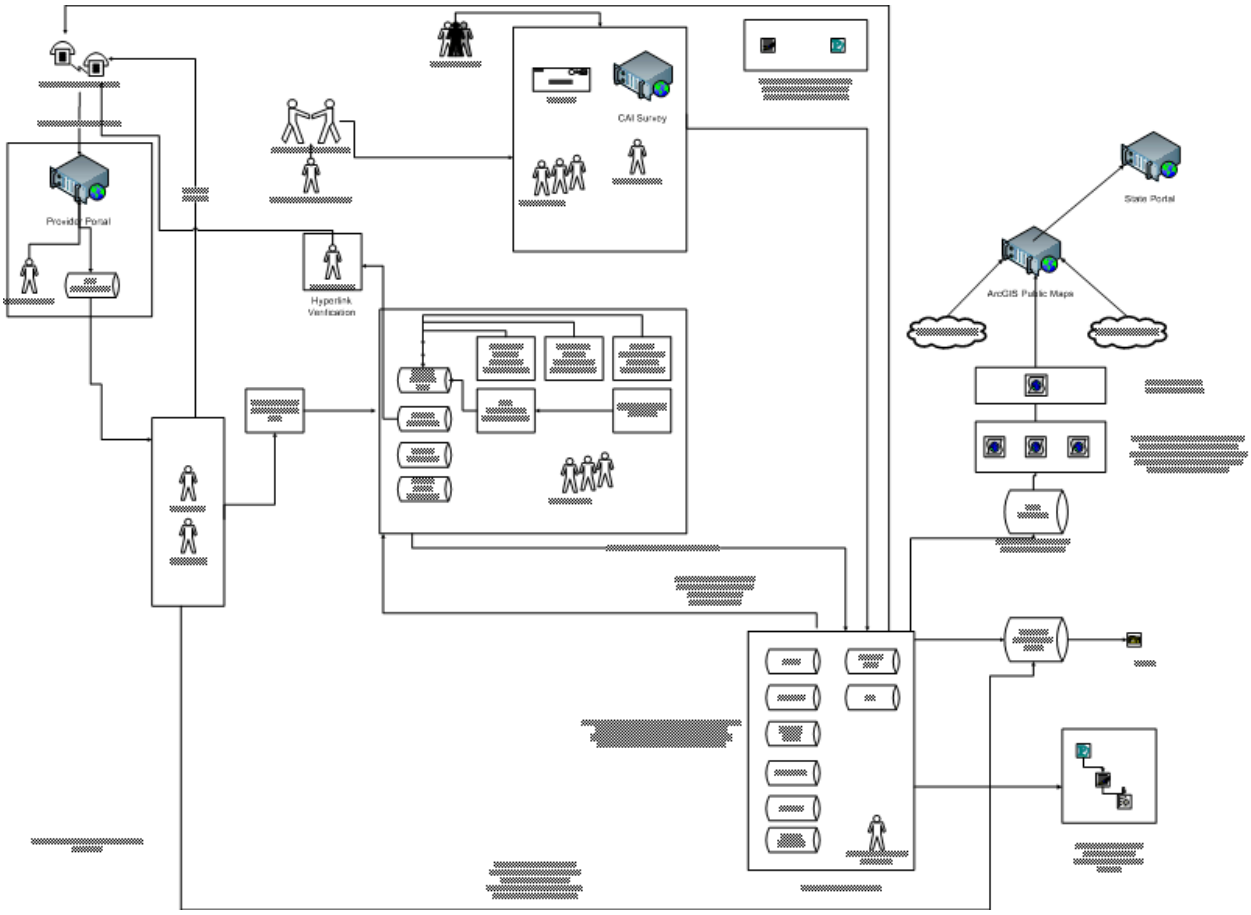


Figure 1—SBI Data Development Business Process Diagram

Data Production Methods

As raw data were received from the provider community, attention turned to normalizing the disparate submission formats[5]. The team considered each submission with respect to the following criteria. These criteria are important because they perform the basis for our verification and quality assurance

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

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

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

Deriving Broadband Coverage Information

Broadband Coverage[6] was normalized into four formats:

1. Coverage in Census Blocks (2010) of 2.00 or less square miles
2. Covered Street Segments (2010) in Census Blocks greater than 2 square miles[7]
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 geo-referenced data
4. Geoprocessing the geo-referenced response
5. Segregating the submission into the correct NOFA-compliant submission formats.
6. Apply appropriate source metadata[8]

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

[7] 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.

[8] 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.

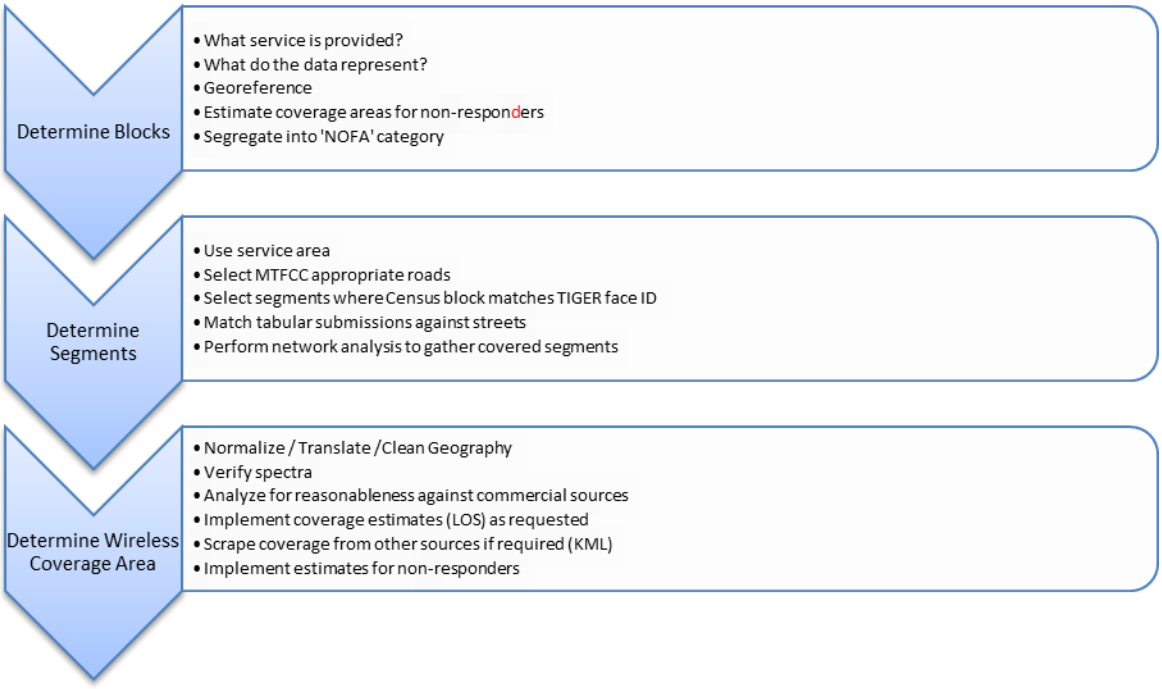


Figure 2-Broadband Coverage Process

Impact of Program Change

There were several important program changes that impacted how Broadband coverage was developed and submitted to NTIA in Round 4.

Census Conversion

The first and most obvious change in submission 4 was the conversion to a Census 2010 coverage baseline. This impacted all wireline providers, the data submitted, the appearance of the mapped information and the baseline coverage metric comparisons against prior submissions.

Release of the June 30 Grantee guidance document, allowed LinkAMERICA to communicate this change with providers. LinkAMERICA provided by FTP access appropriately formatted and sized[9] TIGER 2010 Census blocks and Tiger Road Segments. Given the relatively late release date, we received a mix of responses from Broadband providers. Some easily produced Census 2010 information. Others requested that we do the translation from their supplied blocks and segments. Others requested that we translate their engineering data into appropriate formats. A small number of providers committed to producing Census 2010 data but struggled internally with the conversion in this rapid time frame.

Census 2010 has significantly more Blocks than Census 2000. For the most part there are far more small Census 2010 blocks (less than 2.0 sq mi) than Census 2000. As our team worked through the QA process, this presented a significant challenge in comparing our converted results to prior submissions. We use a block count metric as our first test of consistency across submissions. Since the block count

[9] In Submission 3 we released a technical note describing how we measure Census block area. Although there remains no consensus on this, we used the same process as outlined in the paper.

increased it was hard to distinguish coverage area changes from coverage chances resulting only from a change in Census shapes.

The converse side of this challenge was even more precarious to work through. Because many road segments dropped out due to the covered area now being in a small block area it was difficult to determine how effective our covered segment process was given the fact that many segments naturally dropped out due to changes in Census shapes.

The tendency for large blocks becoming small was not universal. We note in some of our very rural areas of Wyoming and Idaho, small block covered areas become large. This created a contrary situation where small blocks become road segment areas. The image below shows a coverage area change between submission 3 and 4. The covered number of blocks is comparable but the appearance of the coverage is different as a manifestation of the Census change.

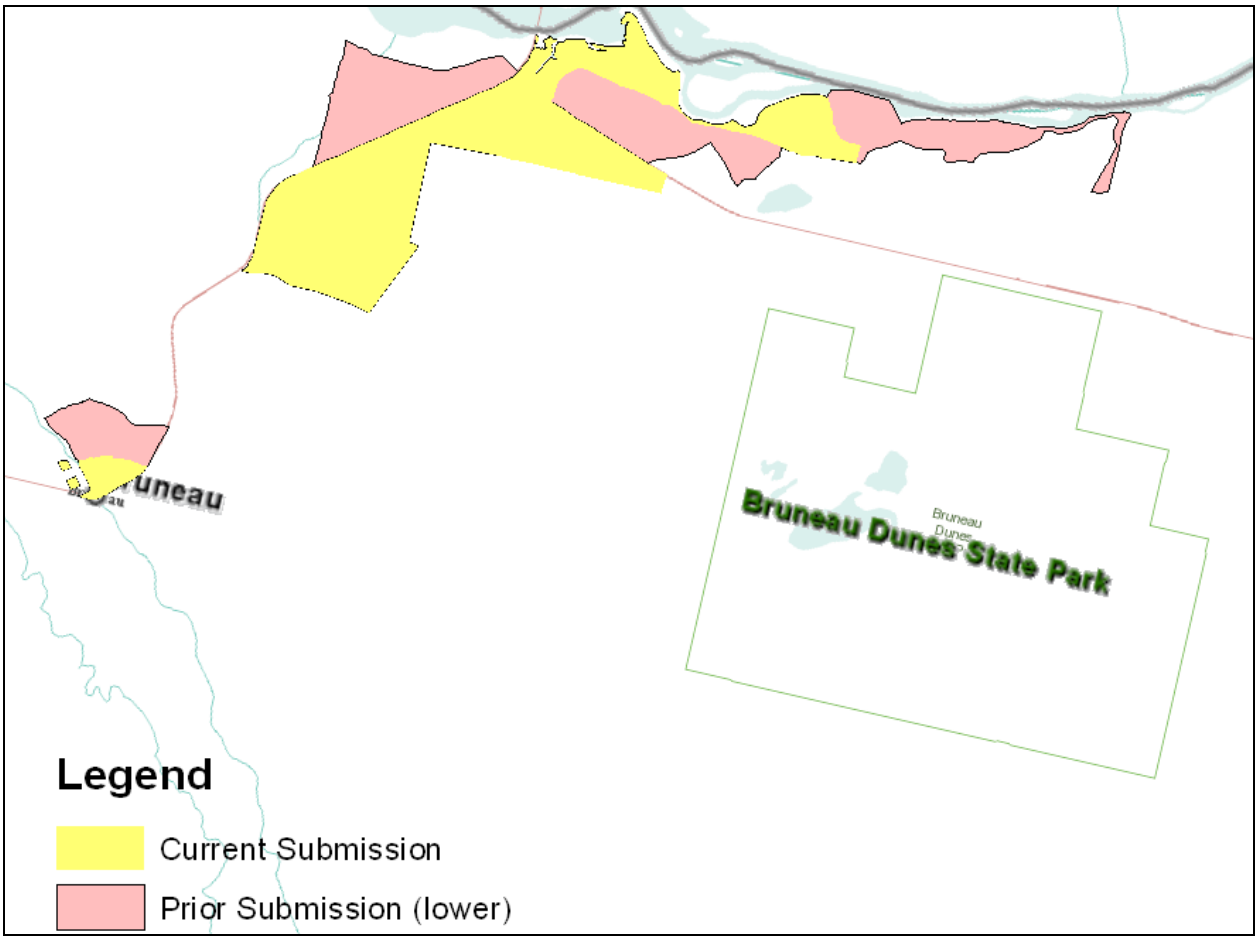


Figure 3--Coverage Change across submissions

This somewhat indeterminate process required our QA analysts to examine a number of submissions in detail. The conclusion was that although the appearance of coverage was significantly different, the underlying engineering data was the same (or very similar) but how the coverage was manifested was a product of the Census conversion.

Census Conversion Practices

Although we had hoped there would be a single process we could follow for all Census conversions our experience has been that it is necessary to be flexible and base the Census conversion process upon the data received.

On a subjective level, we felt the most comfortable converting into Census 2010 where we had facility or demand data to guide the block and segment selection process. In these circumstances we used geoprocessing methods like intersections or network analysis Analyst to make an objective determination. The geoprocessing methods mirrored those discussed in the next section. This was probably the majority of our submitted data.

In circumstances where we were provided Tiger 2010 blocks or segments, we used those as given and performed our standard validation process. Some providers used the TIGER blocks and segments which we supplied them and made their own selections.

Finally, in circumstances where we had either a Census 2000 block list or a geographic file containing Census 2000 geographies and were told there was no coverage change for this submission, we used the Census crosswalk tables[10] to derive a list of candidate blocks. The output of a conversion process is shown below.

[10] See http://www.census.gov/geo/www/2010census/rel_blk.html

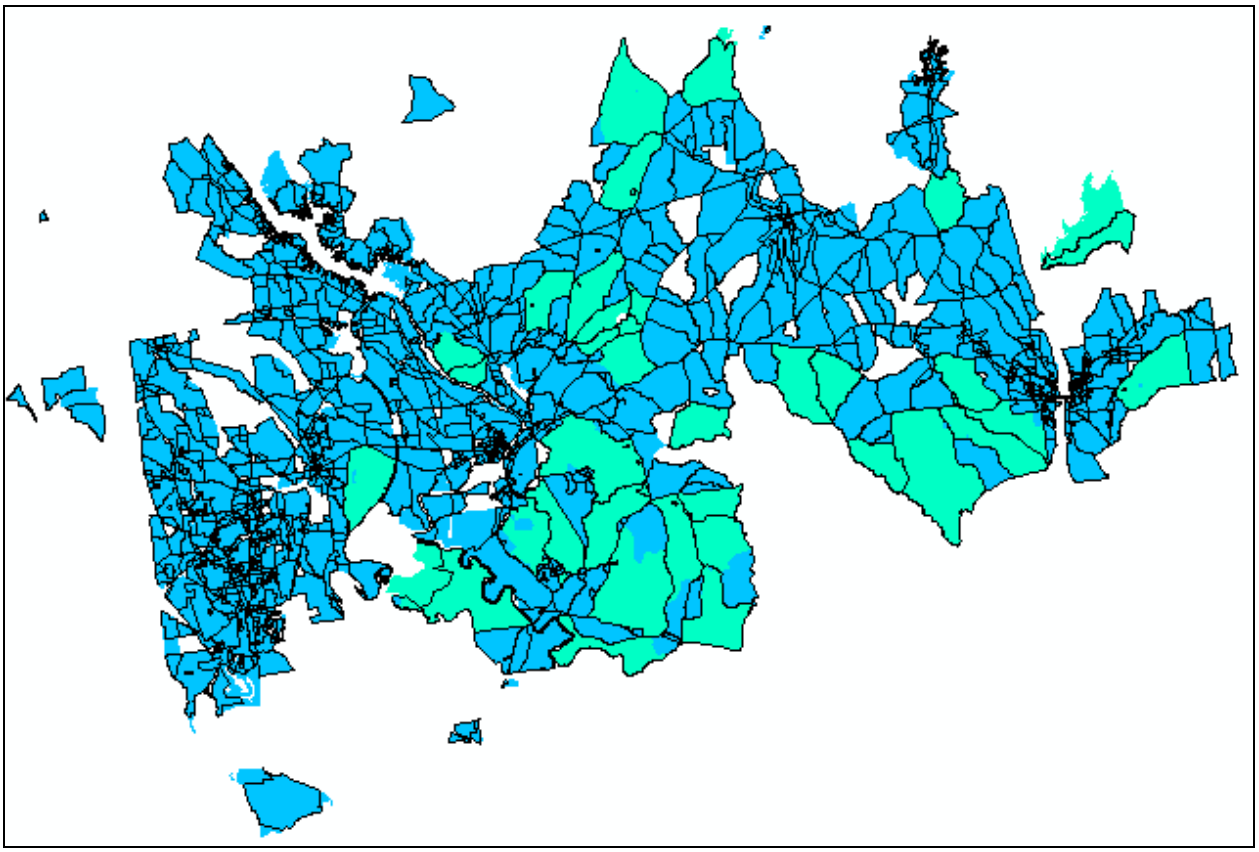


Figure 4—Block Conversion Process, Census 2000 black outline, no fill. Green is 2010 large blocks, so any shading without an outline is 2010 block area not covered in 2000

For the most part it is difficult to discern the impact of a conversion into Census 2010. We don't see vast changes in areas covered. Nonetheless because the block shapes do change the overall coverage area will look different.

As the 2010 data gets pushed into public deliverables, our sense is we will receive questions about the appearance of the new data.

Speed Examination

Given recent concerns about the depiction of speed and what that mapped speed represents, LinkAMERICA invested considerable time requesting detailed information on speed which appeared to be beyond normal speeds for a given Technology of Transmission given the NTIA supplied frequency tables.

Based upon these conversations we learned

A) For a large incumbent telephone provider; the speeds beyond the normal DSL range represent significantly shortened copper loops.

B) For a large national cable provider the intermixing of Docsis 3.0 and non 3.0 systems in a market area is typical and sometimes reflects a circumstance where segments of plant cannot

be upgraded to Docsis 3.0. This variance can be at a level below the Census block. In these cases the maximum advertised speeds remain to represent the market area but the plant variance is typical. This same provider expressed concern with moving reported advertised speeds below the market level.

C) We have a minority of providers who submit a theoretical speed that is unmatched by their web advertising. In these cases we request clarification from the provider on the inconsistency. Our experience has been that providers will modify the speed to be consistent with their web coverage.

Provider Definitions

Within our provider verification process we work to derive a state level provider 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[11].

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[12]. 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 provision 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 subsequent submissions.

[11] We have requested from NTIA information on how provider matching is done within their QA process; beyond the relatively short whitepaper posted with the national map <http://www.broadbandmap.gov/blog/wp-content/uploads/2011/02/DataComparison_Methodology2.pdf>, we have not received any more detailed information on how providers are cross verified between submitted and third party sources at the national level. Our understanding is licensing concerns are holding the release of this information.

[12] We continue to process older submission data looking for information and methods by which we can estimate coverage information. This will be an ongoing process.

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.

After the Grantee conference, LinkAMERICA submitted a paper describing our provider classification system[13]. It is our feeling that understanding the type of provider is essential to appropriate verification methods.

Coverage Geoprocessing Methods

The next section discusses how data were georeferenced and geoprocessed given a particular submission format. We have yet to find a particular method that works across all submissions. Rather we tend to tailor our geoprocessing to meet the specifics of the service provider and data submitted.

In most cases, in Round 4 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.[14]

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.

[13] https://sbdd-granteeworkspace.pbworks.com/w/file/42309493/provider%20ClassificationFINAL.docx

[14] 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.

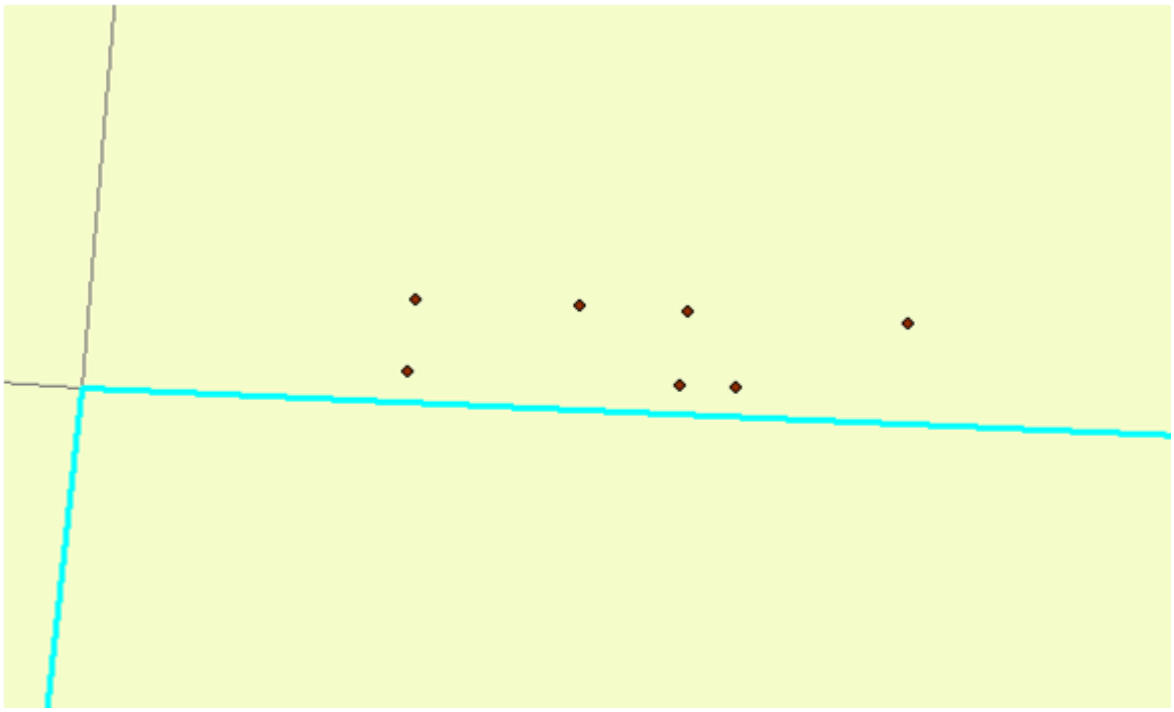


Figure 5-Internal inconsistency in submitted data

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.

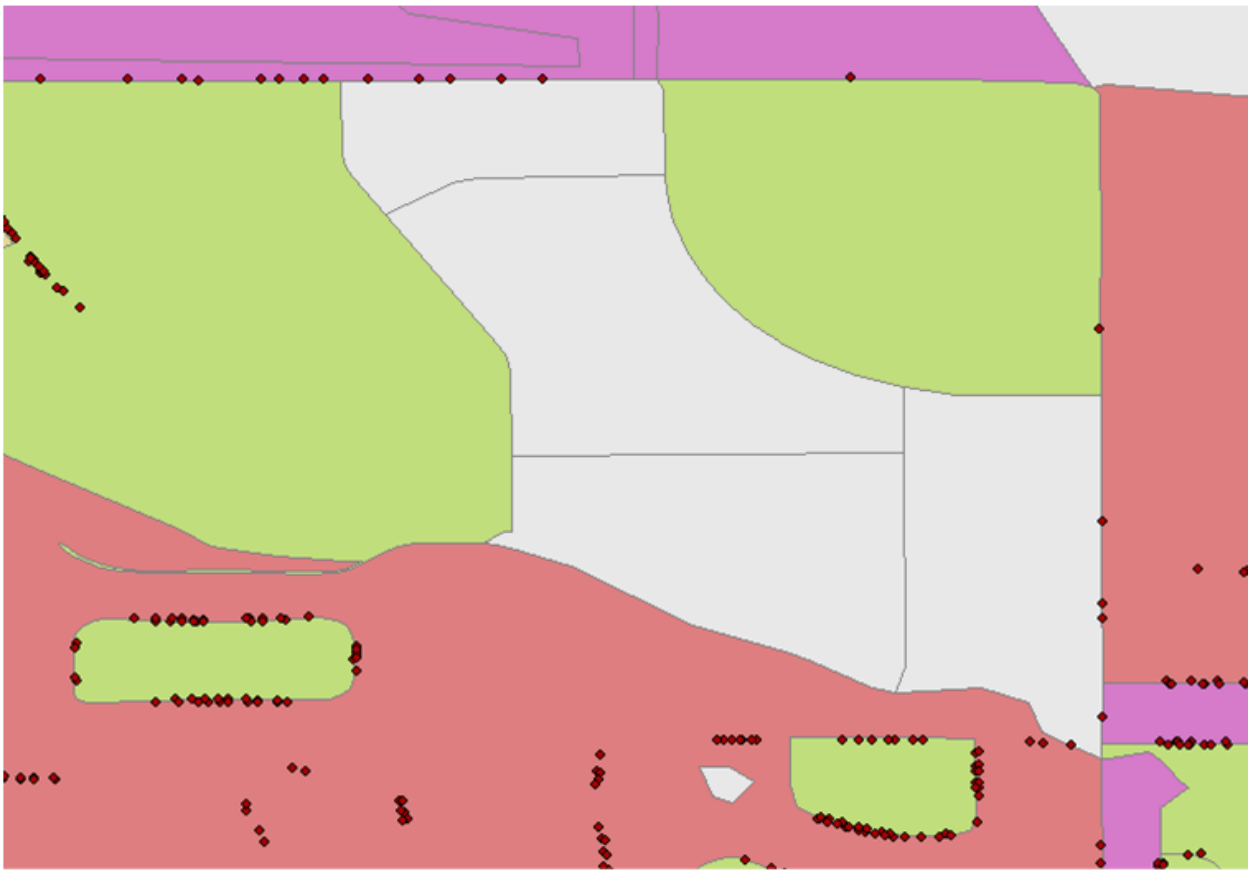


Figure 6-Block Coverage

For this reason, where we were provided address point data and asked to generate covered Census blocks, we elected to use a 200-foot buffer to select Census Blocks that intersect our points.

We are also starting to see a number of providers submit customer data and facility data. Their intent is to allow us to have two primary sources from which to derive the most accurate coverage. In these cases we tend to look for clusters of customers in areas where we see no facility based coverage.

With respect to deriving Block level speed from sub-Block data, we have instituted a business rule where the predominant speed in a Block is the speed we attribute to the Block.

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.

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

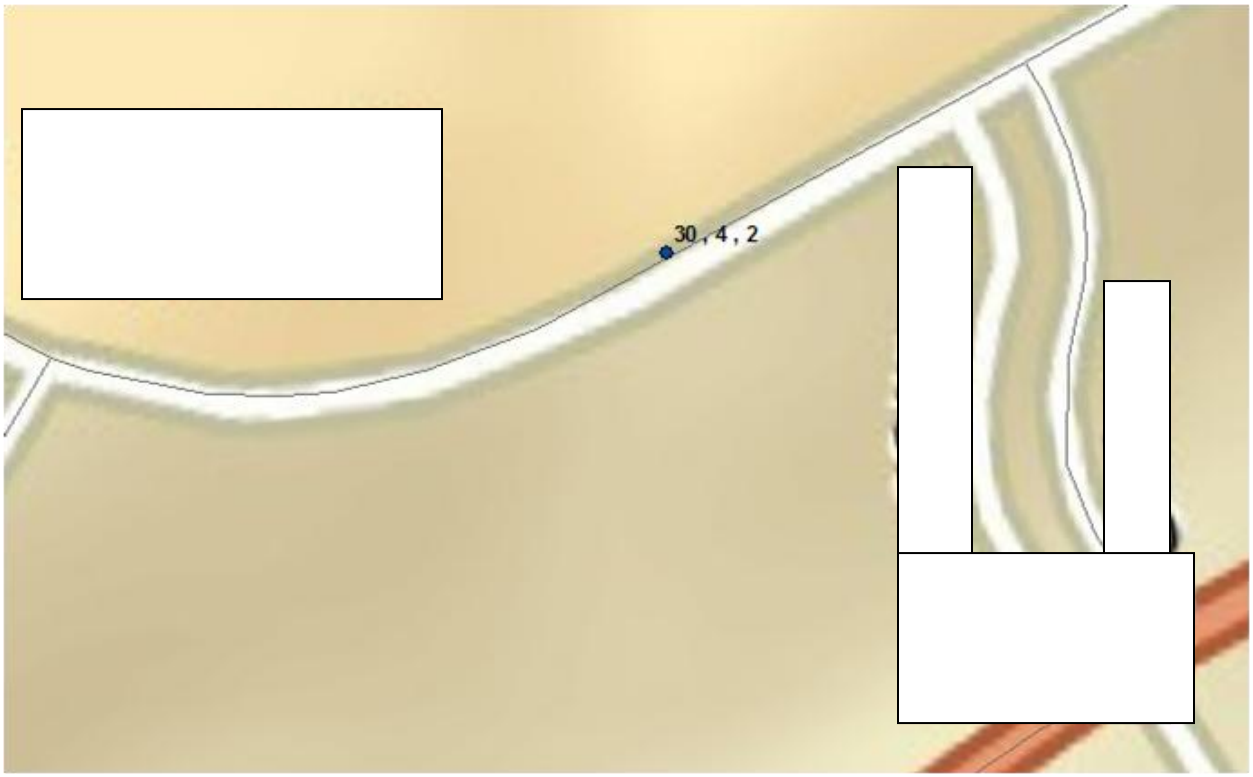


Figure 7-Plant Point level data

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.

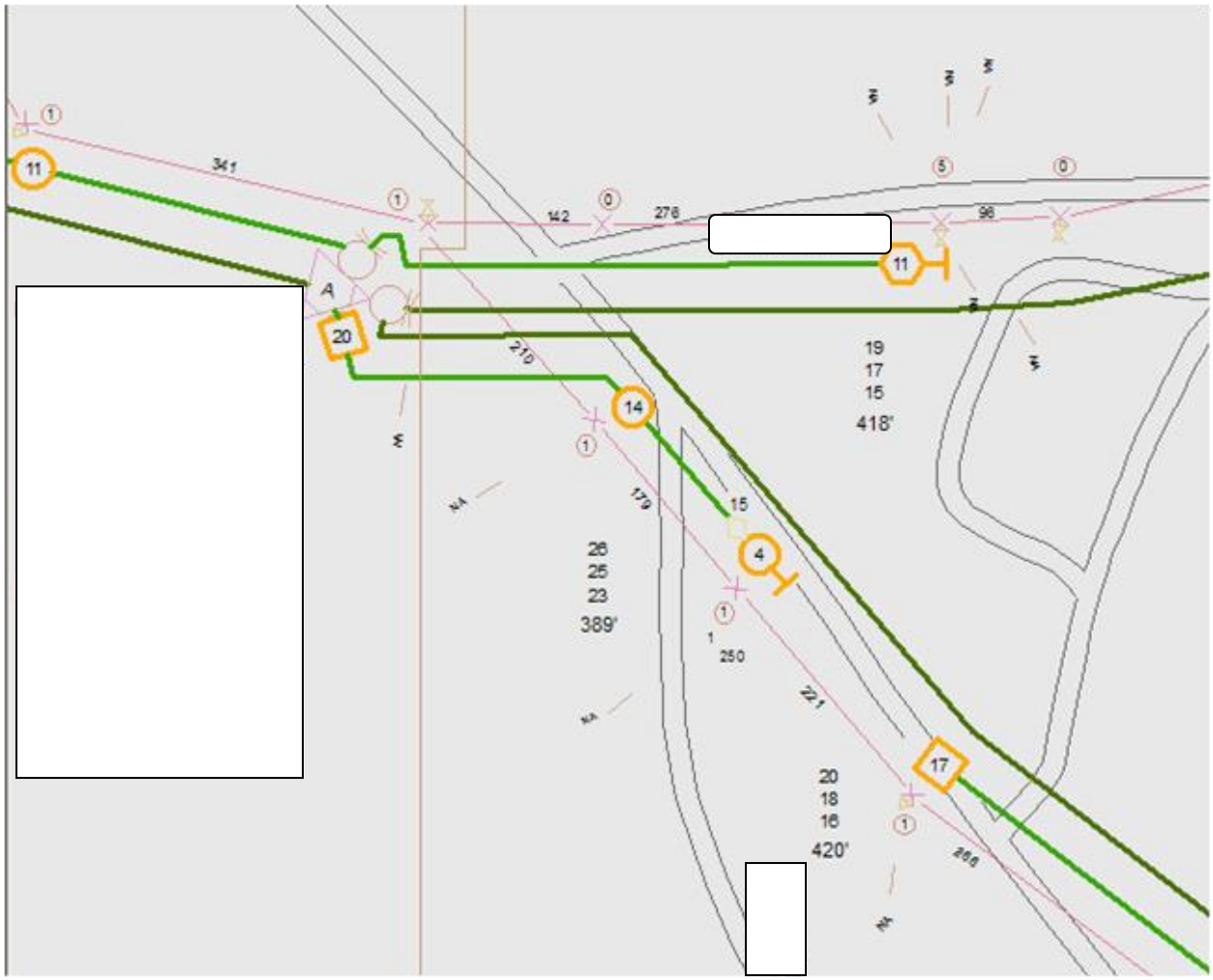


Figure 8-Georeferenced CAD information supplied by Broadband provider

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 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.

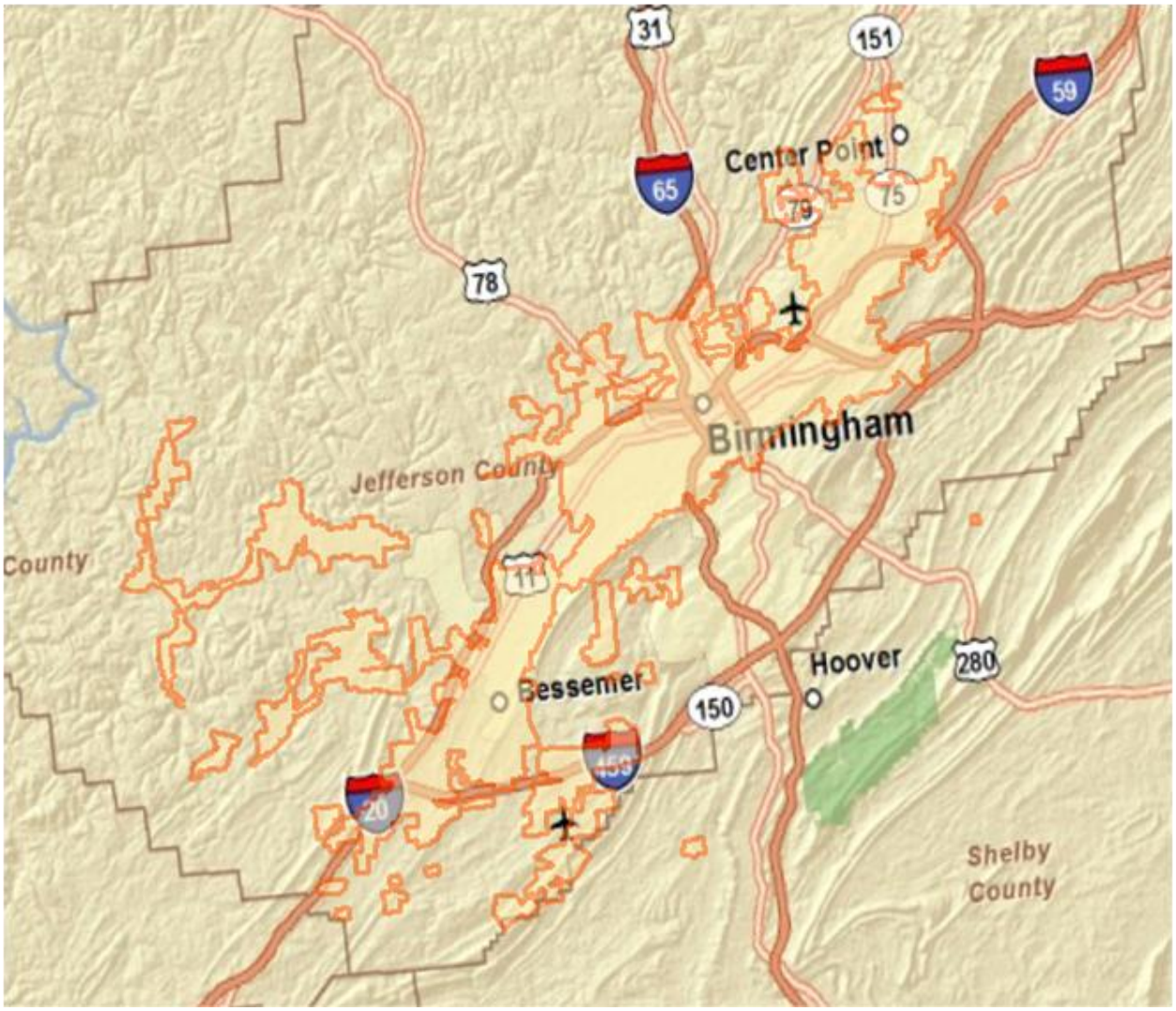


Figure 9-Coverage derived from street segments

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 providers to derive service areas based upon point plant data. In these cases we were given a 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.

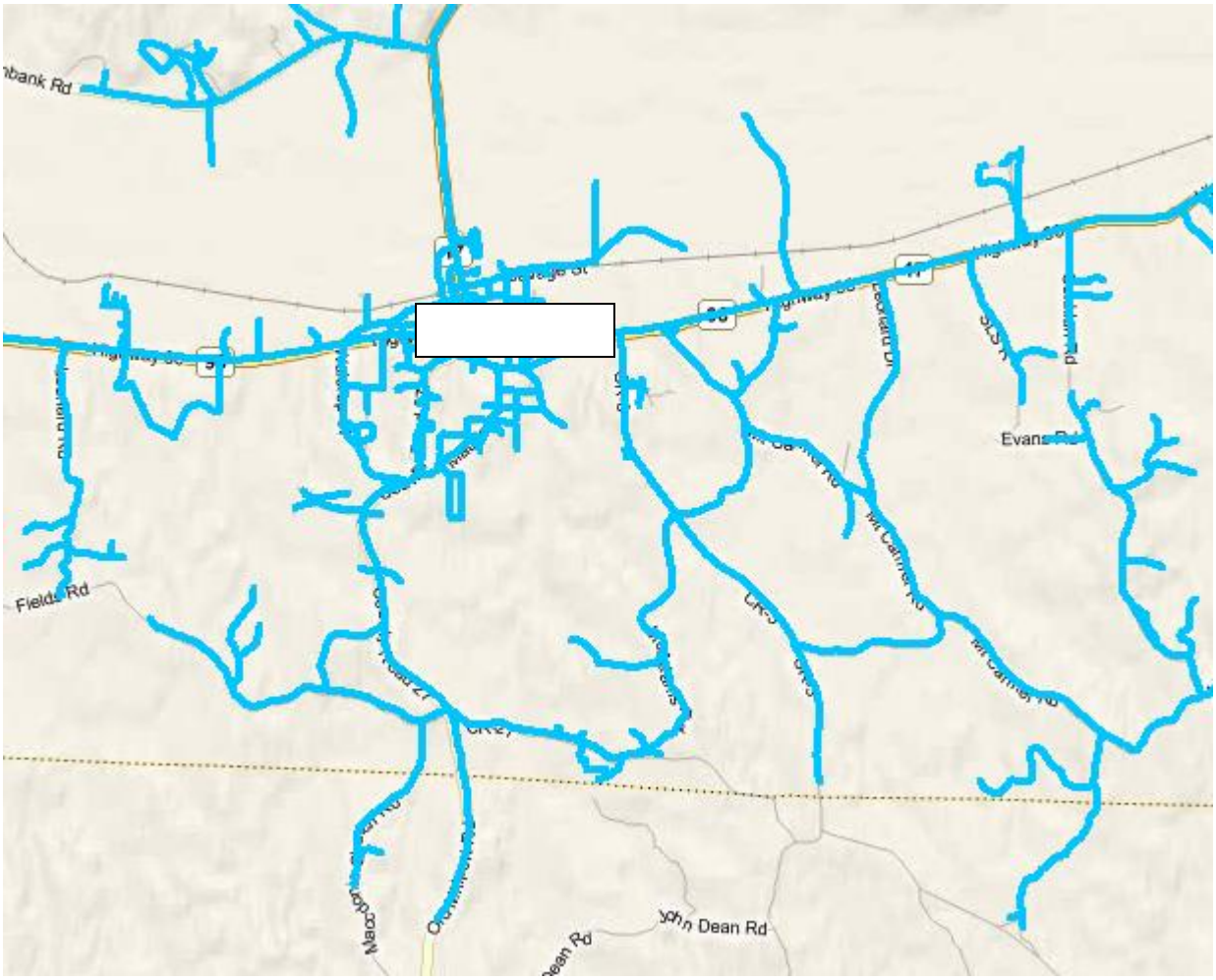


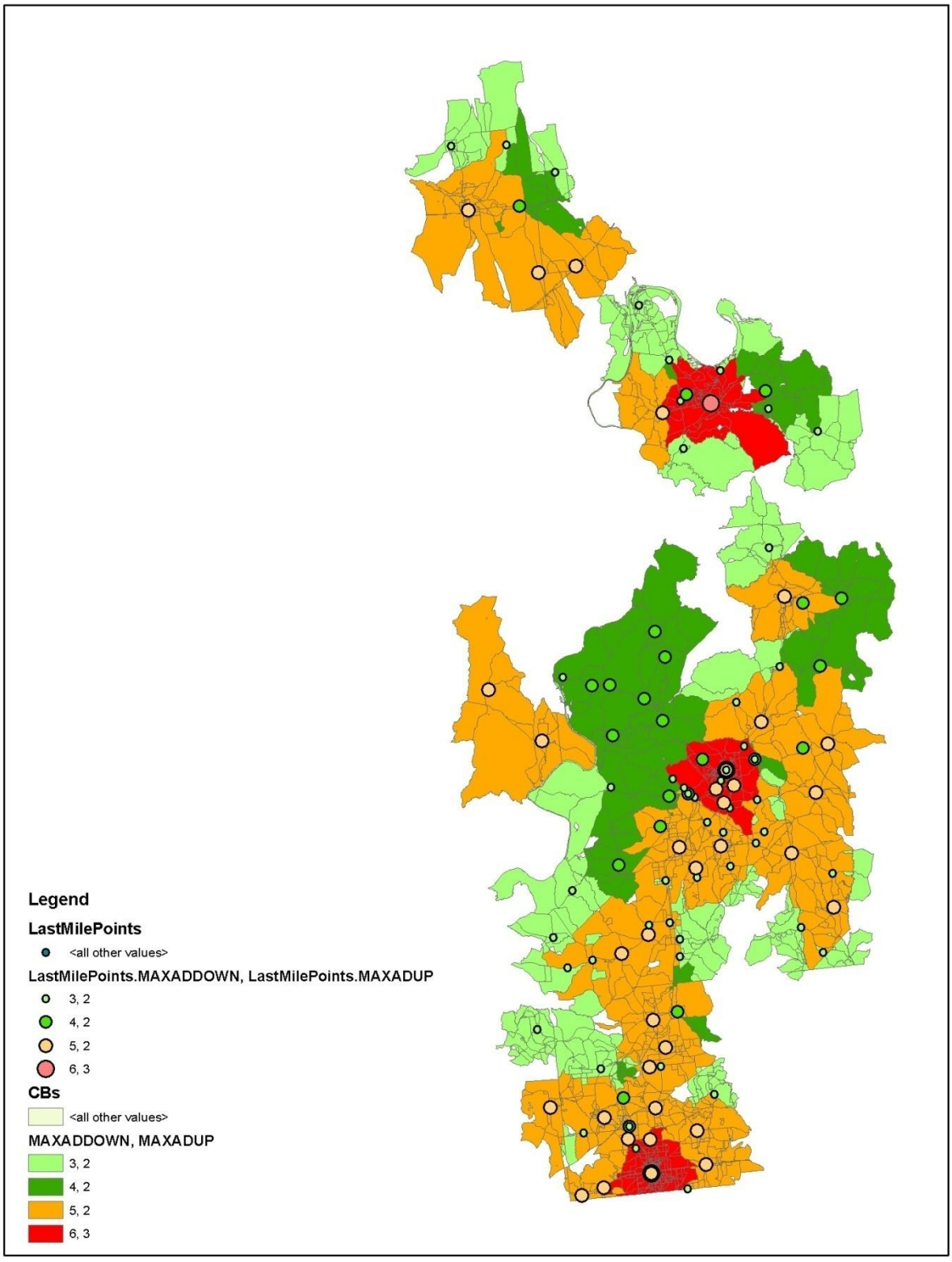
Figure 10-Coverage derived through road paths

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 the current submission -- particularly to improve estimates in areas bordering parks and public lands -- a wider class of TIGER roads was used.[15]

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.

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

Over time we have seen an increase in the number of providers submitting this type of data for our use. Our sense is some providers find plant level data easier to generate and are satisfied with the results of derived coverage.


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

Coverage Derivation Using Polygon/Polyline Serving Areas

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

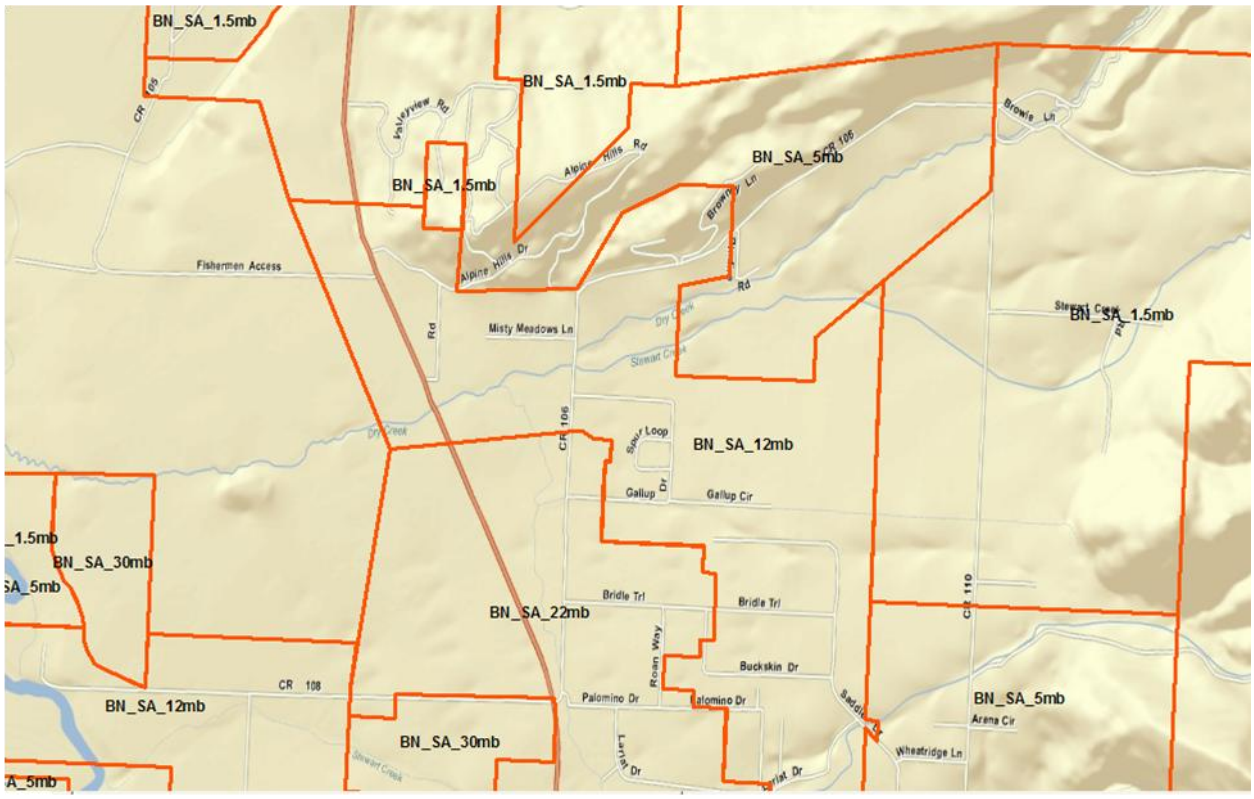


Figure 11-Coverage derived through serving area polygons

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.[16]

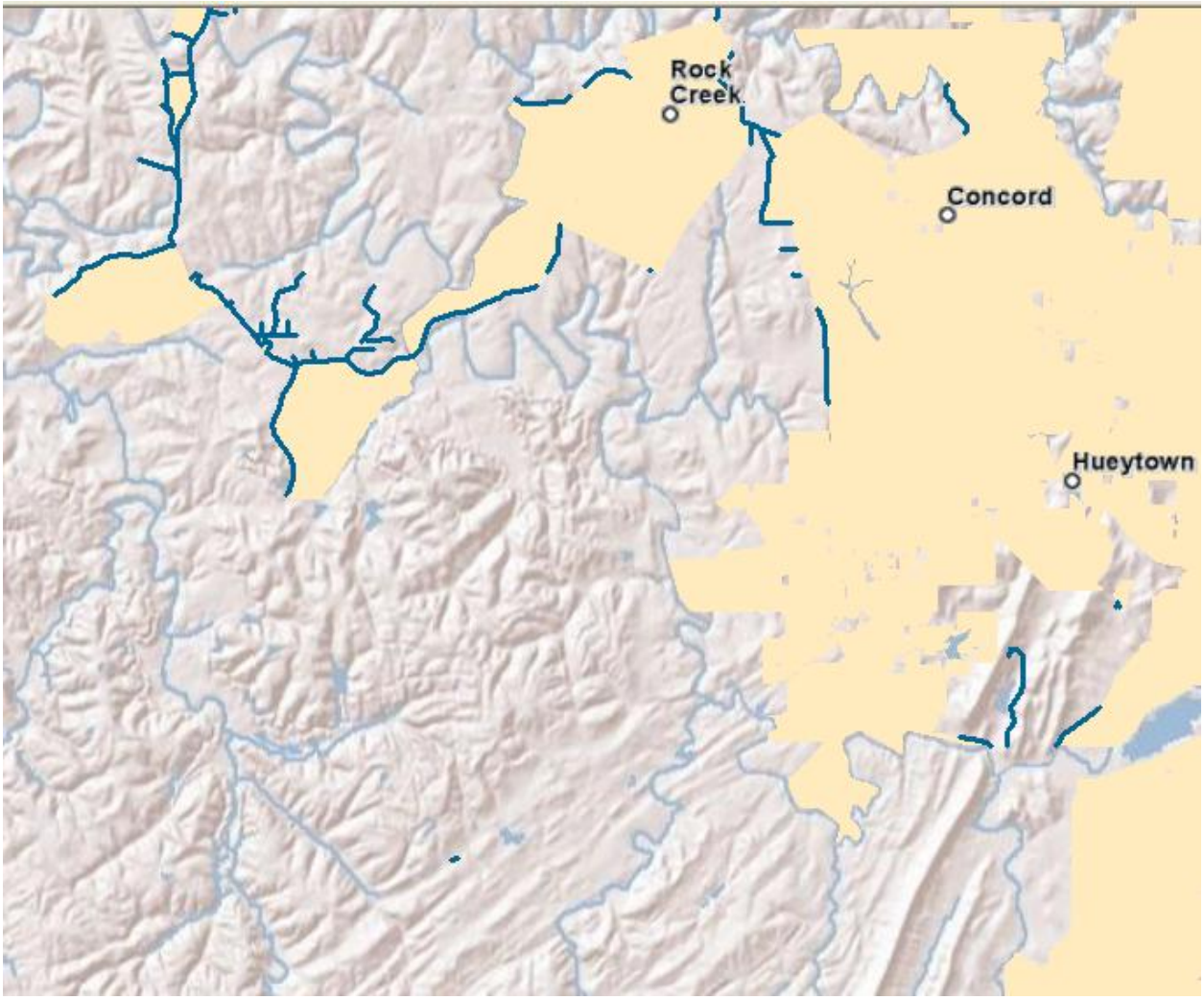


Figure 12-Blue road segments adjacent to peach covered small Blocks

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 2010 basemap. Although there is a trade-off in the accuracy using non-provider submitted segments, we felt it was more important to have a road set that would edgematch our Block features 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.

[16] 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.

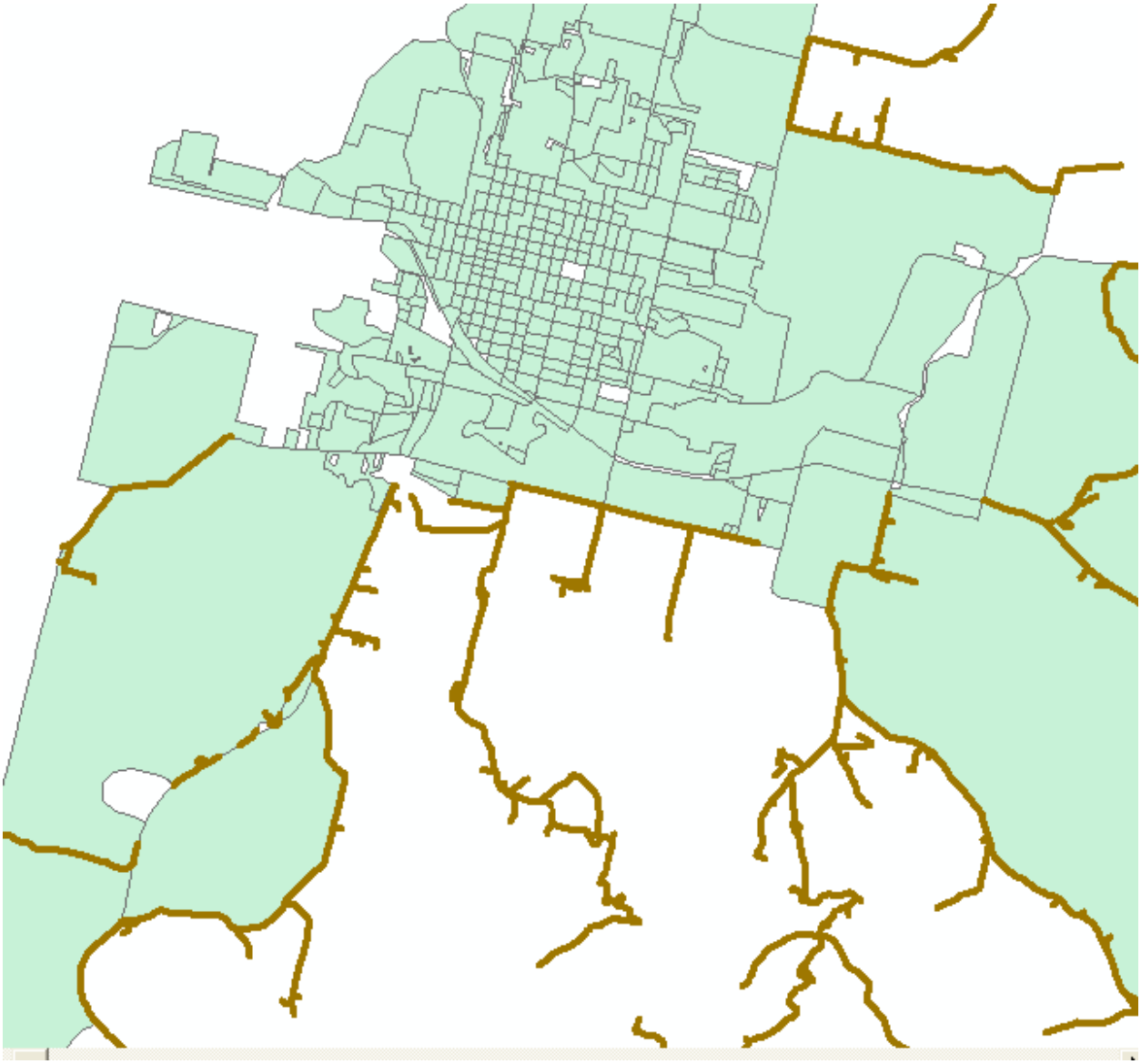


Figure 13-provider Submitted Street Segment Objects. The segments don't edge match the Blocks nor are they continuous.

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

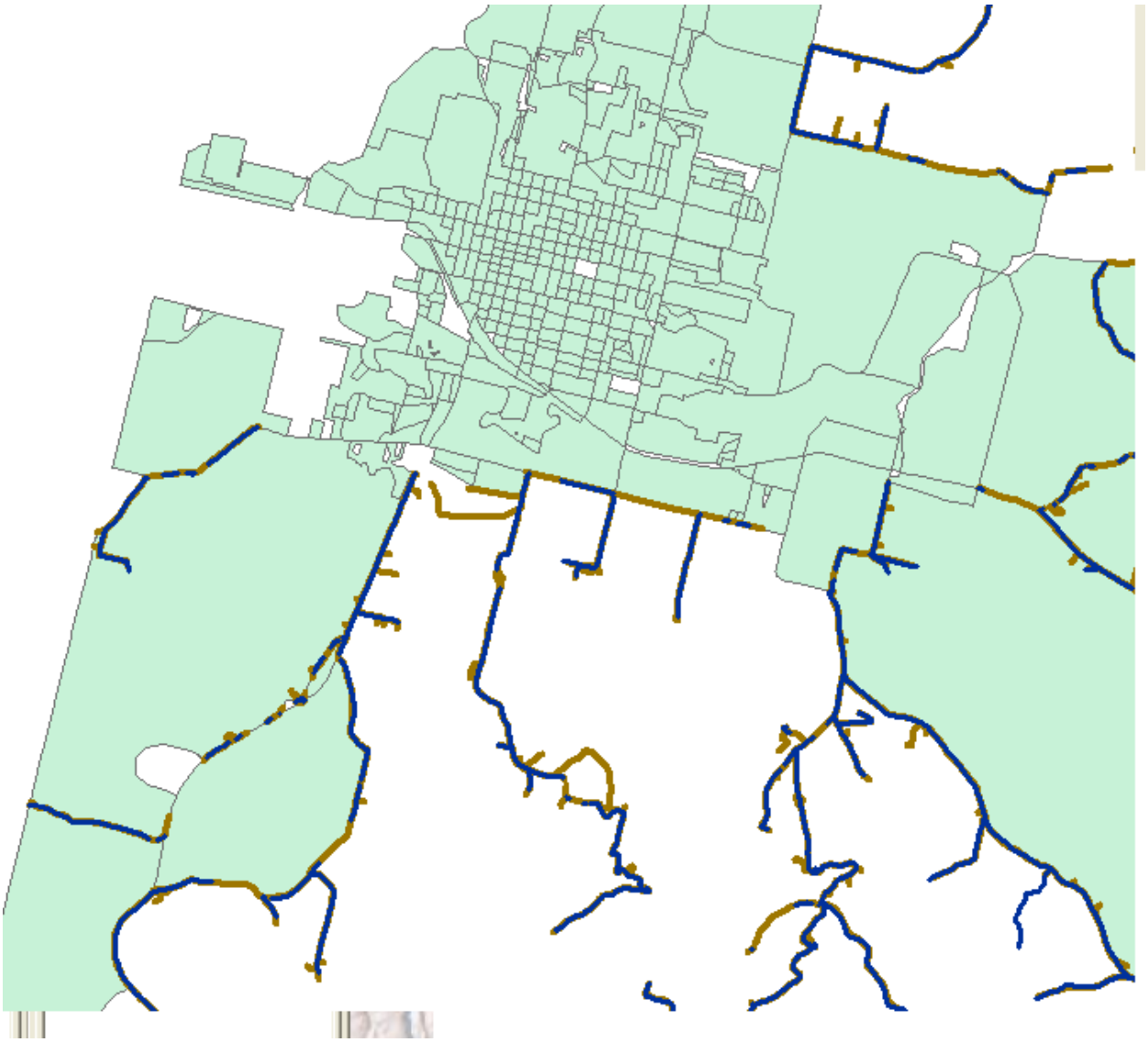


Figure 14-provider submitted segments in gold, selected TIGER in blue—Conflation result; in many cases what was a continuous segment is made discontinuous because even with a distance buffer the TIGER segment doesn't always intersect the provider segment

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 2010 road segments (MTFCC like 'S%') that face (CB = CBLeft2010 or CB = CBRight2010) 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 TIGER roads 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.[17]

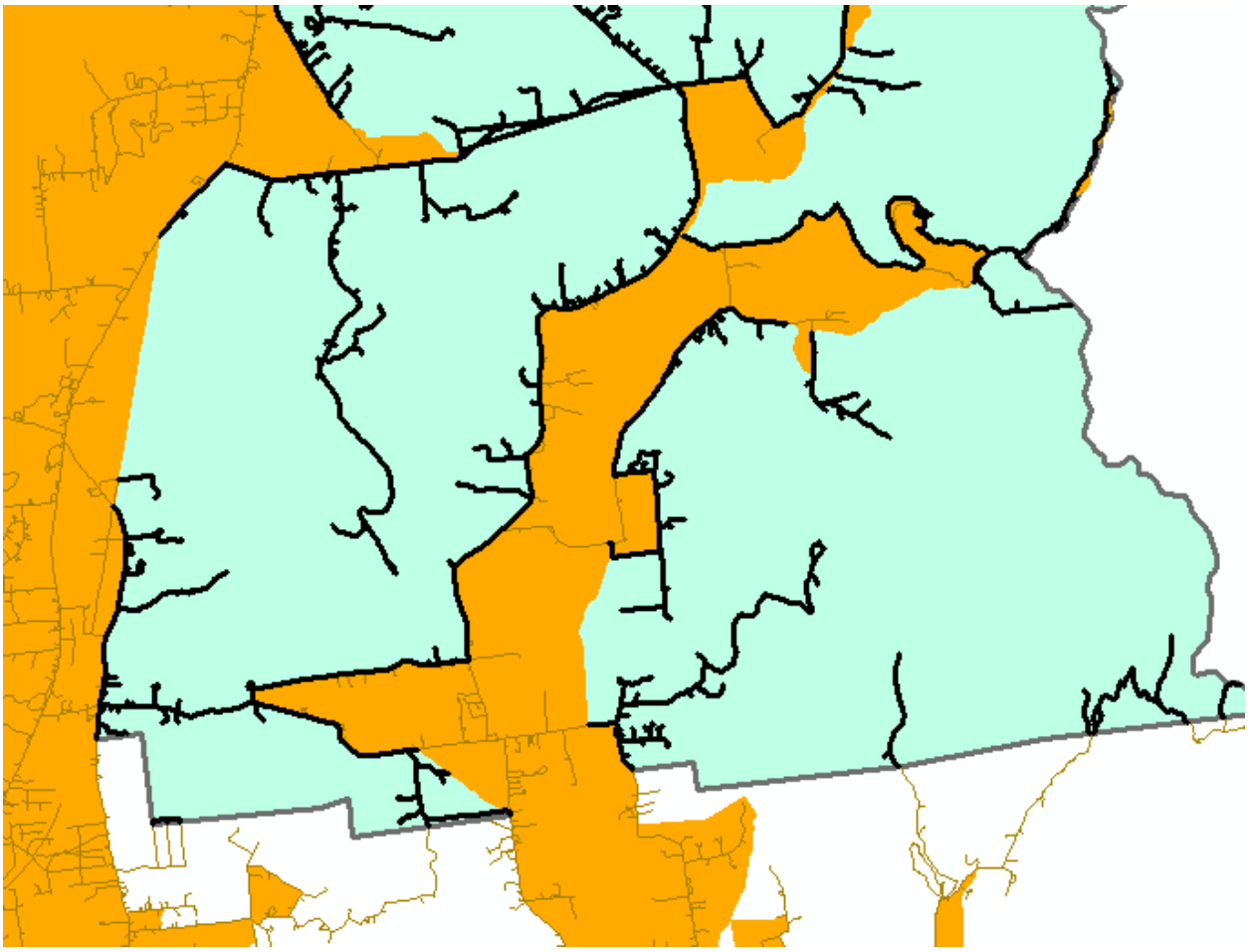


Figure 15-Output of the Segment Process

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.[18]

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

LinkAMERICA continues to make aggressive efforts 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 4 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.

[18] 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 were provided 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 SBI formats. Some of our WISP submissions have taken over 12 hours to normalize into SBI 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[19]. 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. The percentage of excluded data variesacross providers.

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

[19] 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.[20] 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[21].

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[22] 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[23]. 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 derive coverage maps. When these maps were available, we georeferenced them and tried to use the

[20] We will make a second geocoding pass on locations with no longitude or latitude from provider. We typically pick up ~5% from our second geocoding pass. Typically the issue is address quality but also difficulties in geocoding in very rural areas.
[21] We report estimated submissions to NTIA as a non-responsive provider but we have data in the submission for them. This is the reason for datapackage.xls entries which are non responsive but contain submitted data.
[22] 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.
[23] Central Office location was derived from MapInfo ExchangeInfo Professional. Wirecenter boundaries also came from this commercial product.

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[24]. Because much wireless propagation is driven far below the Census Block and much engineering information isn't known (frequency in use, polarization of the signal, coverage pattern of antenna(s), local terrain/land cover) this was the most complicated group to estimate.

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.'[25]

As a general rule, in circumstances where a provider supplies a range of speed attributes, we assign NTIA categories based upon the midpoint of the range. We follow this rule unless we can determine other grantees are handling the same submitted information differently.

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.

[24] 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.

[25] 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.

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 SBI 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 continued to obtain additional connectivity information. For the Spring 2011 collection, the team began a survey process to directly engage these important organizations. As the October 2011 submission represents a transitional phase, much of the CAI effort encompassed getting this dataset stabilized for work outside the LinkAMERICA team.[26]

In the current submission we worked to achieve four goals

1) Modify the source data so as to no longer pass NTIA any connectivity estimates

2) Propagate administrative capabilities in our Community Anchor Verification System (CAVS) systems to the Regional Planning Teams

3) Verify the available connectivity information based upon new survey information

4) Update the Federal record identifiers (NCES codes, etc).

CAI Philosophy

Our work with CAIs is guided by three principles.

First, CAIs are important stakeholders within the planning process. Our goal is to engage participants in regional planning that have strong ties into the 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

[26] LinkAMERICA began transitioning the CAI data collection effort in the state of Alabama to ConnectingALABAMA in Round 3. For Round 4 ConnectingALABAMA assumed full responsibility for the CAI data collection effort in Alabama. To facilitate the reporting process, the ConnectingALABAMA team continued to use the Community Anchor Verification System (CAVS) to store CAI data collected or modified. CostQuest maintained responsibility for the CAI data submission for Alabama for round 4.

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 SBI 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 engaged 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.

Anchor Institution Survey

During the third submission period we designed and developed a simple on-line survey system called CAVS (Community Anchor Verification Survey). The intent of the survey was to both verify received connectivity information and garner additional connectivity information from CAIs. For round 4 we continued the use of the on-line survey process. Although we have found that reaching out to central contacts, for specific institution groups, is the most fruitful way of collecting connectivity data we find value in inviting individual anchor institutions to participate through means of a survey. From our perspective this approach gives the individual institutions an opportunity to become engaged in the broadband planning process. The link for the survey is housed on the Home Page of the website developed for each state, thus providing the added opportunity for responding institutions to learn more about activities in their state.

The survey remains open between collection periods to provide opportunity for the Regional Planning Teams to update information as they engage with the community and to allow responding institutions to update their data as necessary.

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.

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 submissions, we rely on public domain sources of information for the public safety-category . Collecting connectivity data for this group continues to be one of our most significant challenges. 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.[27] 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," We also had difficulty obtaining both the upstream and downstream channel capacities. In most instances, when it was logical to do so, we made the speeds symmetrical, but this is an assumption on our part. If a site records bandwidth across several services (eg. video and data), we record the total bandwidth to give a picture of available site bandwidth. We are also working to standardize our response to NTIA in circumstances where an entity shares a Broadband connection among a campus which is fiber fed. In this case we use the total campus bandwidth and use the primary campus Internet connection.

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 ZIP5.

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 is the site 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.

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

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.")[28]

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).

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 NOFA 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[29]. 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.[30] 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

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

[29] 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}"

[30] 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.

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

Finally with respect to mobile Broadband services, we note several trends.

First LinkAMERICA used the NTIA supplied frequency tables to report speeds consistent with other grantees. In circumstances where a provider supplied a range of experienced speeds, we used the portion of the range consistent with the most frequently reported Grantee value.

Second where a provider reports multiple frequency bands in use but doesn't distinguish these bands by submitted SHP file, we submit identical geometries but attribute one geometry to each submitted spectrum value.

Third we are seeing a trend toward increasing Broadband speed. As of this writing, there is not consistency across providers in how they attribute the advertised 4G speed values. In other words, for some providers 4G means advertised speed categories increase. For other providers, the speed value did not change.

Verification

Almost by definition, data verification is an ongoing and evolving process. Clearly, with each new data submission there will be a validation process at hand and at the same time, our team continues to expand and improve the efficiency and effectiveness our data verification routines. Consistent with 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.
- 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. . 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.

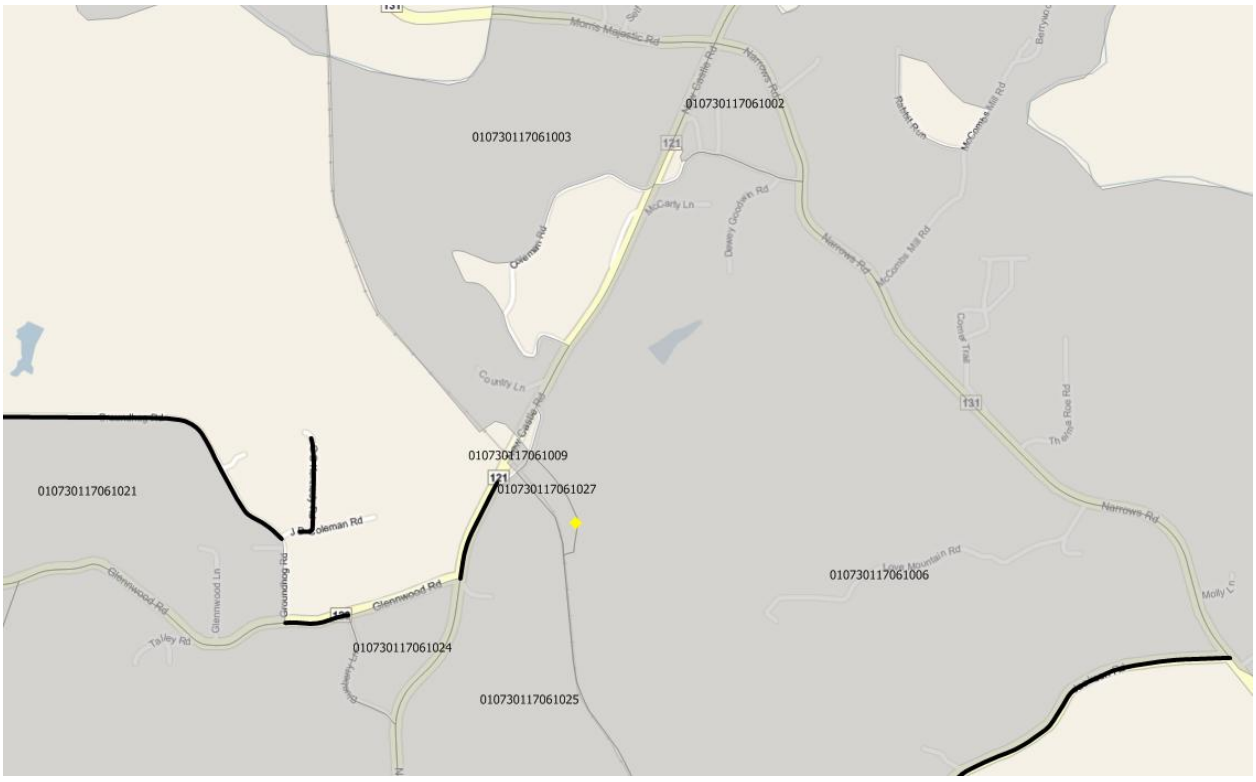


Figure 16--Sub block variation

Because the SBI 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.

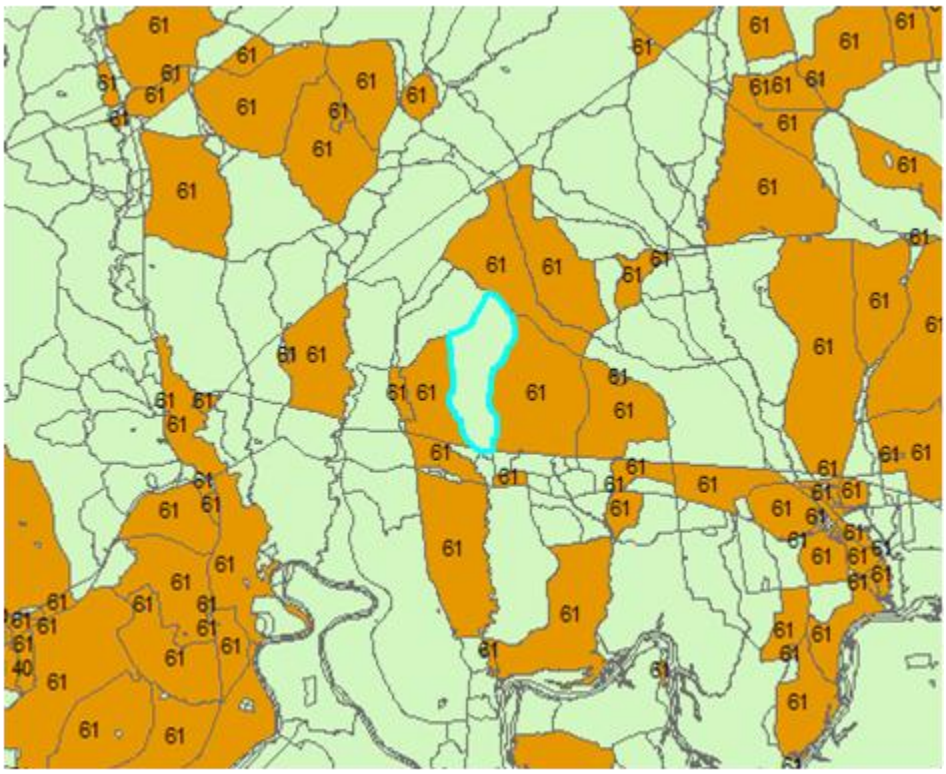


Figure 17--Dispersion in Submitted Data

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 shows where an interpolation process could require 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.

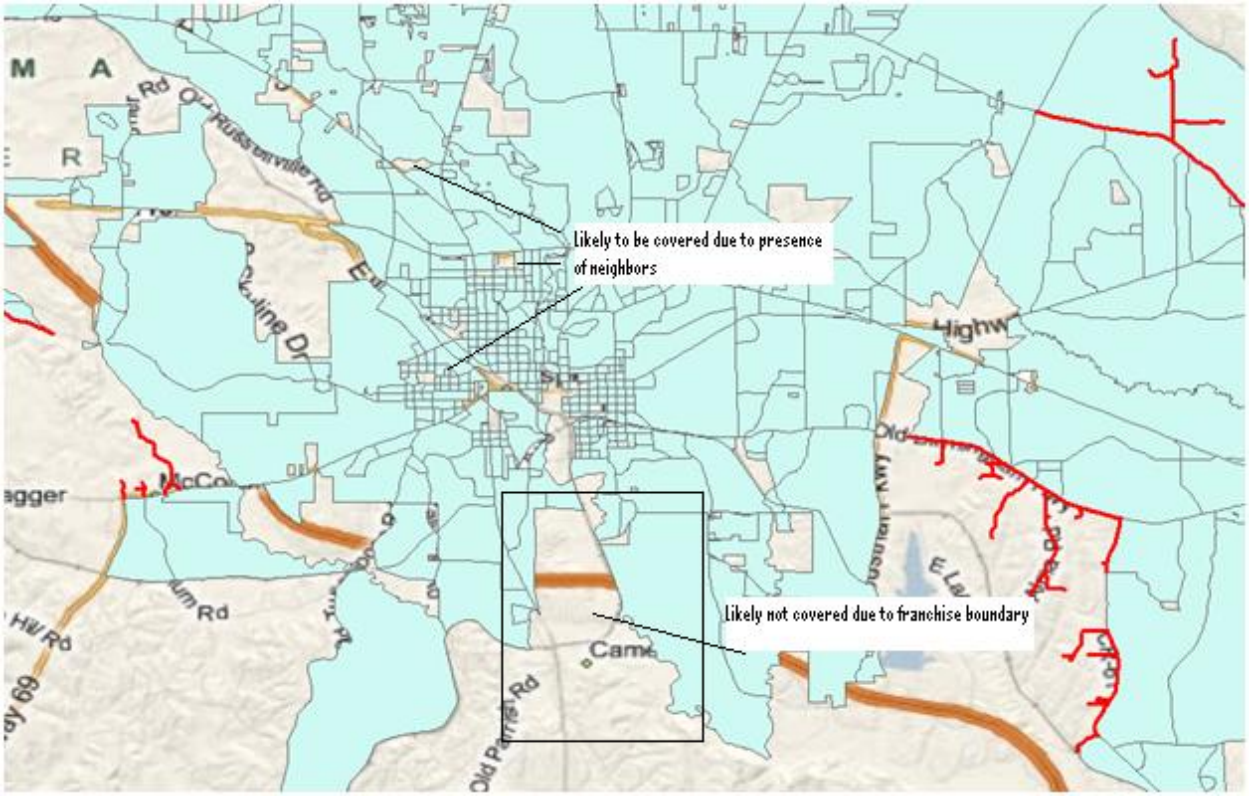


Figure 18-Where do you stop interpolating?

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.

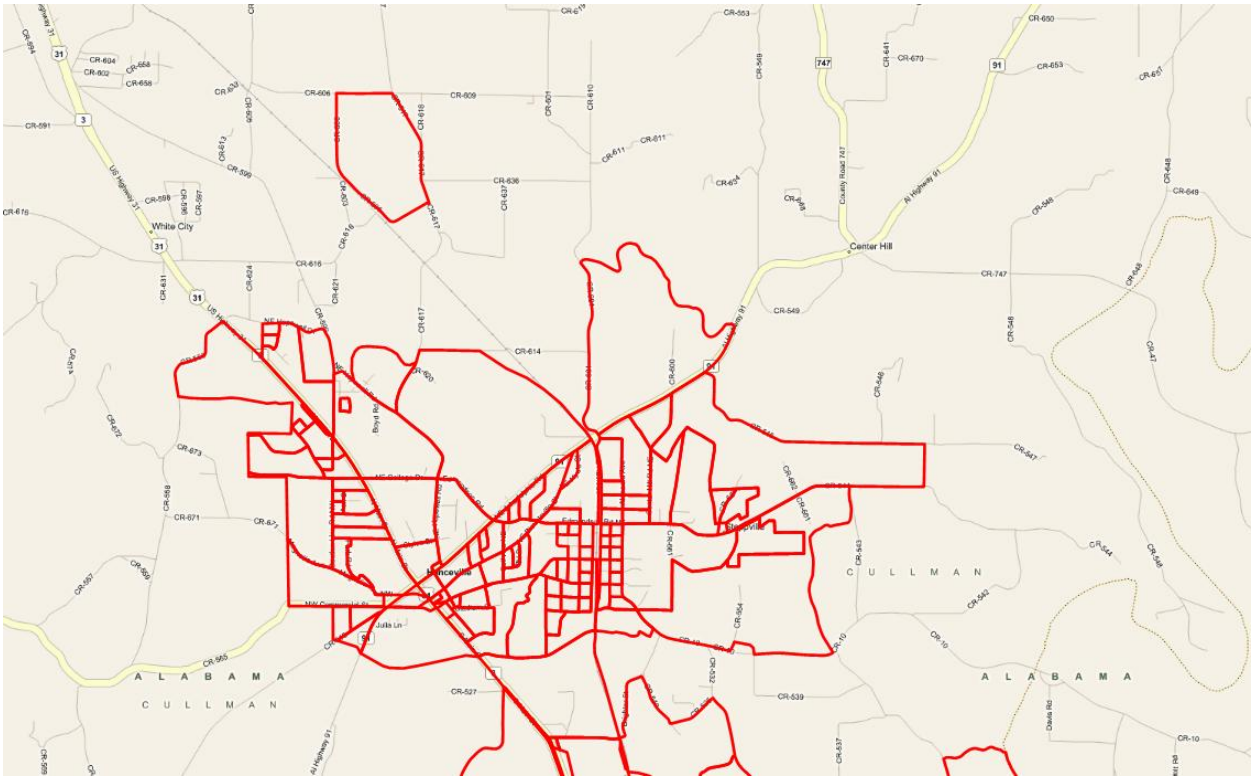


Figure 19-Dispersion in covered Blocks

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 third round, they received both a tabular data summary and mapped output[31]. Prior to releasing the "check maps" to providers, we had a team of analysts visually inspect each provider's coverage area. The focus on this QC effort has been to identify and flag suspect Blocks. After this in-house review, we solicited a second level of feedback from providers and received a number of requested changes and corrections used in the development of the October, 2011 round 4 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. We also verify the submitted speeds against the typical speed ranges in the NTIA frequency tables. If we note a value outside of typical range, we ask the provider for clarification.

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.

[31] For the verification of round 3 data, we submitted both PDF and KMZ (Google Earth) format check maps. Some providers prefer to work with the Google format as it supports easier modification. Others continue to submit marked up PDFs.

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

With 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.

Online Display of Consumer Feedback

We have completed development of a consumer feedback layer for our online maps.

The intent of the new layer is to show viewers the feedback of other map viewers. We anticipate the feedback layer will go live when the Round 4 data is posted on our state maps. We expect this to be prior to the end of October, 2011.

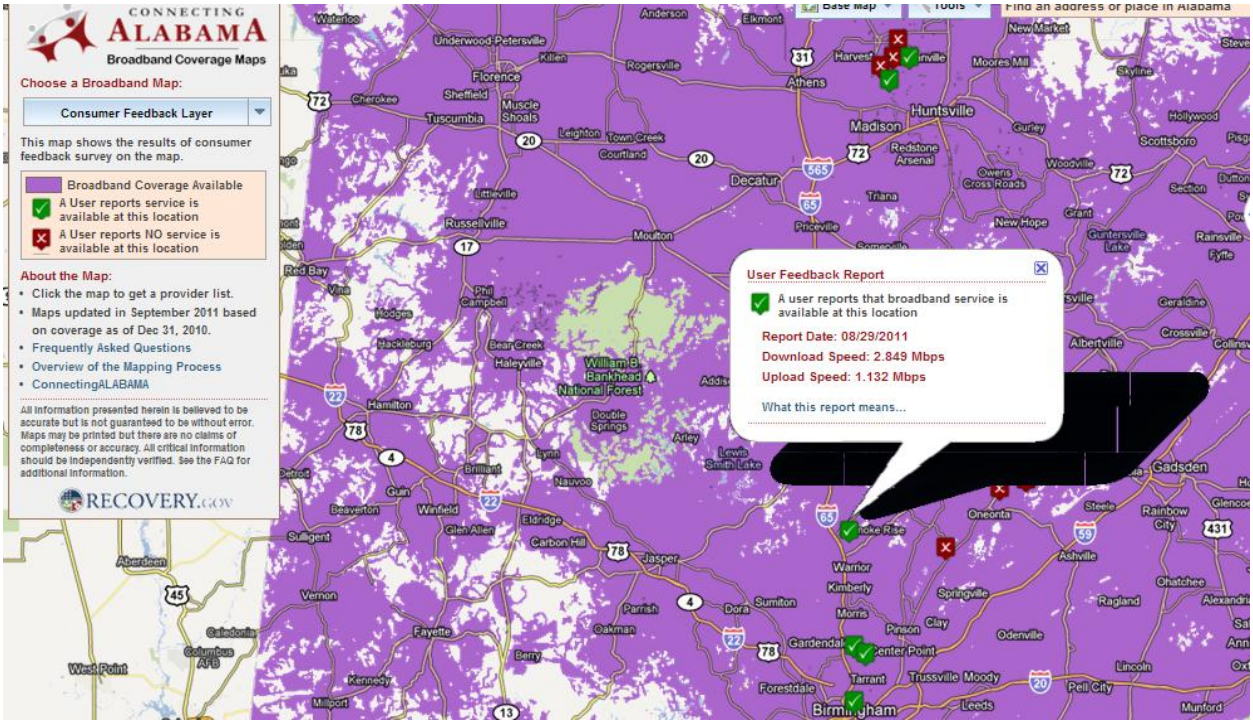
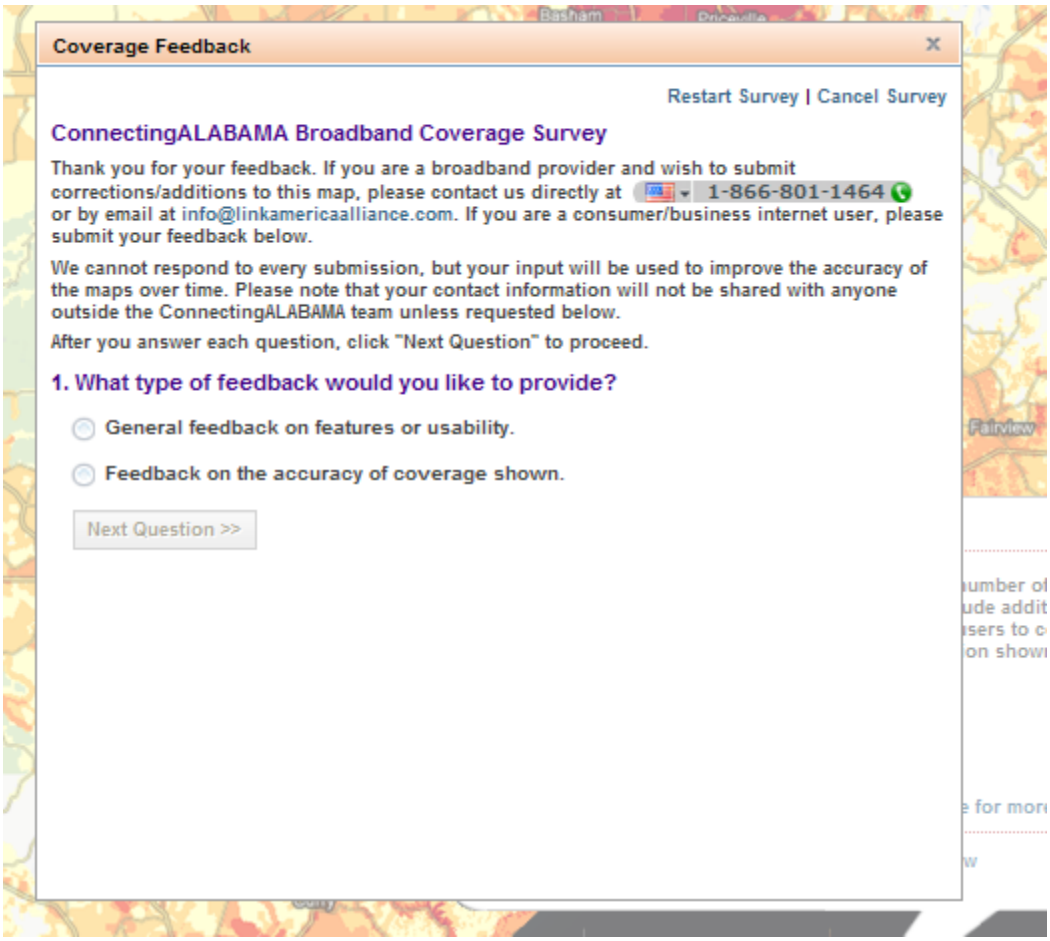


Figure 21--Consumer Feedback Layer

To gather feedback, we use a survey wizard which asks the end users to categorize their concerns. The survey went through several iterations of design and usability testing. Our experience has been unless we get a way to constrain the user feedback into manageable categories, it becomes very difficult to act upon.



As mentioned by other Grantees we struggle with how to use all of the feedback we receive. The qualified data points seem to fall below a volume in which we can infer significant modifications to the map data. Nevertheless, we believe it is important to gather structure and display the feedback to support project transparency.

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 SBI data model. As an example, consider the figure below.

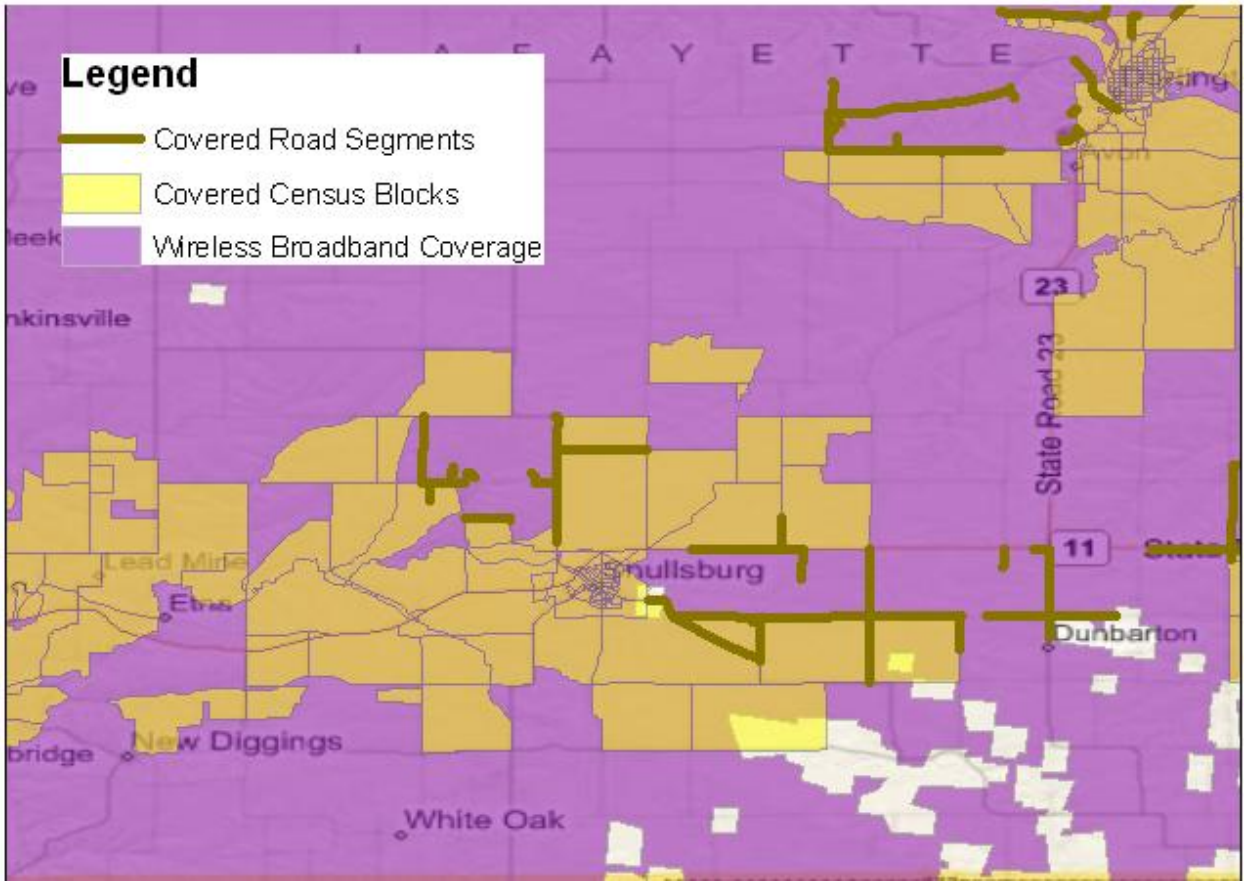


Figure 22--Multi Network Coverage portrayal

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). Some 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 Carrier of Last Resort (COLR) Obligations

Some 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' from road centerline. 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 specific 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.

CAI Survey Fatigue

We are beginning to note an increase in survey fatigue among CAIs. Sometimes, as part of a direct survey process an end user will tell us how unhappy they are with the repeated Broadband survey efforts. Within several states BTOP grants are in effect that also survey Community Anchor Institutions.. As stated earlier we will defer to other Grantees when there are overlapping survey efforts.

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

We receive a variety of Census data. Some were able to supply 2010 Census blocks. Others continued to provide Census 2000. Managing this set of heterogeneous inputs has proved to be a challenge.

There seem to be several methods by which providers are calculating the Census block 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[32]. 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 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.

[32] 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.

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[33]. 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 Issus

Non-Uniform Submission Standards

It is clear among providers that there isn’t a consistent method used to derive Broadband coverage. Some providers appear to be use 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 was used to georeference plant (is it the carrier’s roadbase?). This leads to uncertainty regarding the input data scale or accuracy relative to other base layers. Although we may be trading off absolute

[33] 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 ProviderType 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.

accuracy, our standard has been to conflate submitted data to TIGER 2010 Blocks and TIGER 2010 roads. We perform our verification against this conflated data product.

Temporal

We are unsure of how well the data are temporally consistent. Some providers gave us their best effort to control to June 30, 2011. 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 note 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 a requested geographic deliverable also increases the complexity of the effort.

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 Street and Zip5 attributes(Service Address), if the value is unavailable it is filled with N/A. was not available, we have inserted 'N/A
6. As with submission three, there remains a tension between the Data Model, Data Model Default Values and the Python Validation Script. As an example the data model allows a NULL for the Maximum Advertised speeds in a Census block record. A default 'zz' is available for this condition as well but zz will fail the validation script. In the case where we have data which is missing Maximum Advertised Speeds, we are holding that data back to prevent downstream validation problems.
7. We have a significant amount of VDSL, ADSL 2 and ADSL 2+ coverage categorized into the xADSL category. This introduces a variance in speed availability as some providers are using VDSL, shortened loops and/or pair bonding to increase speed over 10 Mbps.

8. We have left in the data Middle Mile locations with above grade elevations that appear to be unreasonable, given review of orthoimagery. This seems to be confusion between above grade request and above sea level readings.
9. All fGDB have passed validation except in cases where attributed speeds did not agree with domains associated with technology of transmission (eg Upstream Speed of 2 with ADSL). We have modified the Python script to allow for conditions in the CAI table in which default data model values are disallowed in the Python submission script.
10. 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. We have verified on provider websites that the reported speeds are available in the area but these speeds will fall out of the NTIA frequency table analysis.
11. We have a small number of providers who serve an area with both a residential and business speed tier. In cases where we cannot distinguish which speed tier offering to use, we use the lower of the speed tiers.
12. Per NTIA request we have modified the manner in which we handle Wireless coverage polygons. If a Provider submits a single geometry but specifies multiple spectrum codes in use in that polygon, we duplicate the polygon for each spectrum code. In other words the geographic object is identical but the attribute data for the object is unique.
13. In point level data submissions (Service Address and CAI) we note points that are spatially coincident. With respect to Service Address points our thought is these represent multi-unit dwellings or businesses but we don't have enough address detail to determine if these are multi-unit structures or duplicated customers. Because we cannot determine the reason for the duplication we leave spatially coincident records in our submission. We also leave in our CAI submission points which may be the same physical structure but have slight variations in addressing.
14. In point level middle mile data, we are finding a variance in the quality of the geocoded longitude and latitude returned. Given the data received we are unsure if this is an issue where the plant address is difficult to geocode or if the longitude and latitude provided to us is different than what would be returned in geocoding.
15. We note two important issues in our datapackage.xls. First the number of records in the provider tab will not sum up to the total record count. This is due to the requested grouping within the Excel table.. Second for estimated broadband coverage, we internally mark that coverage as an estimate but the provider is described as non-responsive within the datapackage.xls.
16. We made one modification to the NTIA supplied verification script. For the CAI layer we The query to check the TRANSTECH field now includes: "AND TRANSTECH <> -9999"

Appendix Two

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

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

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

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

	There are no restrictions on distribution of the data by users