New Jersey Broadband Mapping Project:

Methodology Report on Data Integration and Validation Procedures For September 2011 Submission

September 30, 2011

Grantee:
New Jersey Office of Information Technology
200/300 Riverview Plaza
PO Box 212
Trenton, NJ 08625

Contact:
Shelley Bates
shelley.bates@oit.state.nj.us
609-633-9605

Contractor:
Telcordia Technologies, Inc.
1 Telcordia Drive
Piscataway, NJ 08854

Contact:
John R. Wullert, II
jwullert@telcordia.com
732-699-2687

Data Processing: Collection, Reception, Loading, Validation

This document presents a description of the process used by the New Jersey Office of Information Technology (OIT) and Telcordia Technologies to collect, receive, load, validate and verify broadband availability and usage data submitted to us by wireless and wireline service providers, CAIs, and other sources and organizations for the State of New Jersey. Individual provider data reports attached hereto provide details on each provider's submission and explain how the policies presented in this document were applied to the data. The CAI summary report, also attached, provides details on the CAI data processing. This report also describes some of the complexities and challenges we have encountered to date in this project.

1 Structure of this Report

This methodology report consists of the following

- Section 2 summarizes our outreach efforts to collect data
  - This section also describes some of the challenges in determining what service providers are in and out of scope for this work and our approach to service provider categorization, in addition to summarizing our efforts to engage CAI constituencies
- Section 3 provides an overview of our process for Service Provider Data Reception
- Section 4 provides an overview of our process for Service Provider Data Loading
- Section 5 provides an overview of our process for Data Validation
  - This section includes a table of business rules and how they were implemented.
- Section 6 provides additional details on two issues related to Geometry
- Appendix A: NJ Provider Data Reports
  - This appendix concatenates 36 files in Microsoft Word format, one file for each provider whose data was included in the submission. Each report provides a narrative describing the steps involved in collecting, verifying, loading, and validating the provider data, including a log of the interactions with the provider.
- Appendix B: CAI Processing Report
  - This is a summary of the details of the CAI processing for this submission.

2 Data Outreach

2.1 Provider Data Outreach

Telcordia and OIT have conducted further outreach this summer to identify additional potential resellers as well as providers not previously participating. We have used web searches and email and telephone contact to investigate the status of these organizations with respect to the NOFA definitions and the goals of this project. OIT will negotiate NDAs with those providers who request them. Providers are given instructions on data requirements, including how to submit via our custom-designed Web site found at http://connectingnj.state.nj.us/.

Most providers were willing to participate, although several have expressed concerns about the burdens of the data collection process. One provider – Hotwire Communications – previously declined to devote any effort to submitting data. The large national providers clearly have processes in place to collect and submit data, while the small local providers require greater assistance. Telcordia offers assistance where possible, allowing providers to submit whatever data they have available in any convenient format. This increases the complexity of the data collection and processing operations, but enables greater coverage of providers. As examples, some smaller wireline providers simply submitted a list of addresses where they offer service and some small cable operators submitted the names of the municipalities they cover.

- In this round we have submitted availability data from 35 facilities-based providers plus one reseller, including five organizations that are new to our program this round (Clearwire, Level 2, NetCarrier Telecom, and Network Billing Systems are new providers and New Edge Networks dba Earthlink Business is a new reseller). We also continued to include the three satellite providers whose data we first submitted in April (i.e., Hughes, Starband and Wildblue). Our initial company list at the onset of this project came from FCC aggregate Form-477 data that we receive under the Form-477 sharing arrangement. We have been subsequently working to expand this list by screening other potential providers and resellers. In addition we have been tracking the evolution of the provider community over

time – this includes mergers and acquisitions among organizations as well as organizations that expand their region of operation and go in or out of business.

- There are numerous web-based sources and aggregators that provide information on potential broadband service providers and resellers. As just one example, the Broadband Internet Directory (http://broadband.theispguide.com/) is a consumer website that lists broadband offerings and plans. Other examples are www.dslone.net/nj, www.globalspec.com, www.broadbandinfo.com, etc. We periodically review these sources to identify organizations that may be relevant for this program.
- The broadband industry is dynamic with mergers and acquisitions taking place regularly. We track the consolidation of entities, among other reasons, because the availability data may not reflect the larger organization for some time after the closure of the transaction. Some of the transactions we are currently tracking include: PaeTec acquisition of Cavalier; CenturyLink acquisitions of Qwest and Savvis; MegaPath acquisitions of Covad and Speakeasy; Earthlink acquisition of One Communications; Appia Services acquisition of Voxitas; etc.
- On the reseller front, there is a wide range of entities that fit rather differently into this program, ranging from resellers like New Edge Networks whose data is included in this submission to MetTel who does not maintain engineering data about customer service technologies. We would also like to note that Global Crossing was very responsive to our outreach. As a facilities-based provider who does not meet the 7-10 service provision window, however, they do not meet the NOFA definition.

2.2 Service Provider Classification

We have classified Service Providers into the four categories as follows:

Type 1 = Broadband
These are broadband providers that meet the NOFA definition of a facilities-based provider with a 7-10 service provision time frame.

Type 2 = Reseller
These are broadband providers who do not meet the NOFA definition of a facilities-based provider because they resell facilities that belong to another service provider.

Type 3= Other
These are broadband providers who are known not to be of Type 1 or Type 2. Typically this is either because they cannot meet the 7-10 day service provision time frame or because their service architecture is complex and is neither facilities-based nor a reseller.

Type 4 = N/A
We are not currently using Type 4.

Since it is only Type 1 providers who are squarely in scope for this program, these are the only ones for whom we have ensured that the NDA, provider_ind and submit_ind columns in the service_provider_info spreadsheet are completed. Our rationale for this is the following -- we would not want to categorize a non-Type-1 organization as “will not provide data” or “non-responsive” under provider_ind, as this may appear pejorative.

In our ongoing efforts to reach out to the full set of broadband service providers in New Jersey, we work to identify potential providers and screen them to determine if they are providing or reselling broadband services in the state. We maintain a commented list of those organizations that we have determined not to be New Jersey broadband providers or resellers and of those organizations that remain under investigation. Some of these organizations are no longer active business concerns; some are no longer independent organizations, but have been acquired by other entities; some offer or resell broadband service in other locations but not in New Jersey; some are companies that provide engineering or consulting support around broadband, but do not provide or resell service; and some are firms for which further interaction is needed to definitely determine their situation. Service Providers

During the summer we initiated additional outreach to try and identify potential broadband service providers or resellers and then determine their categorization. The impetus for this effort was the program’s expansion of focus to include resellers and the additional service provider types. Our efforts resulted in the categorization of twenty-nine additional organizations:

- Four additional Type 1 service providers whose data is included in this submission: Clearwire, Level 3, NetCarrier Telecom, Network Billing Systems.

- One additional Type 2 service provider whose data is included in this submission: New Edge Networks dba Earthink Business.
- Two additional Type 3 service providers: airBand Communications (fixed wireless provider with service in Philadelphia; they have one location in New Jersey from which they cannot serve additional customers) and Global Crossing (cannot typically meet the 7-10 day service provisioning time frame).
- Sixteen organizations for which we are still in the process of determining their status and role in the industry.
- Six organizations that are neither broadband service providers nor resellers; these firms are summarized in the table below.

Name of Company	URL	Explanation
American Telephone Company LLC	americantelephoneinc.com	Equipment provider
DatNet Communications Group, Inc.	See under lightower.com	Acquired in 2007 by Lightower.
Hickory Tech Corporation	hickorytech.com	Not currently offering service in New Jersey
Towerstream, Inc.	towerstream.com	Not currently offering service in New Jersey.
World Discount Communications Co.	mywdt.com	Provides discount calling cards.
Yipes Holdings, Inc.	Redirects to Reliance Globecom	Yipes was acquired in 2007 by Reliance.

2.3 CAI Data Outreach

Telcordia and OIT used a variety of means to collect Community Anchor institution data. We collected reference data with lists of CAIs of various types in the state and we collected broadband data from individual institutions via our website and from aggregated sources. For healthcare institutions we had previously obtained a reference list of hospitals from the New Jersey Hospital Association and we augmented this with information parsed from the New Jersey Department of Health and Senior Services (NJ HSS) which maintains on-line records of all licensed health care facilities. For K-12 education we augmented our broadband records with information extracted from NJ applications to the federal e-Rate program. For the e-Rate program, we obtained public information on all New Jersey applications from the USAC website. There are five funding categories established in the e-Rate program, plus a Miscellaneous category. We selected applications that requested funding for the Internet Access category. The available information allowed us to identify these schools as having broadband access

For each CAI category, the following table provides the number of records we obtained from the reference source, the number of broadband access records we obtained, the total number of records we submitted to the NTIA and the number of complete records, with verified address information and broadband access information.

CAI Category	Reference Records	Broadband Records	Total Records	Complete Records	Submitted Records
School K-12 (Public)	2603	796 (Web) 478 (eRate)	2598	175	3518
School K-12 (Private)	1260 (NCES)		1267	169	
Libraries	465 (IMLS)	89	472	50	443

CAI Category	Reference Records	Broadband Records	Total Records	Complete Records	Submitted Records
Medical/Healthcare	1107 (NJ-DHHS)	5	1107	5	1106
Public Safety	343 (NJ 911 Comm.)	120	349	104	328
University	158 (NCES IPEDS)	39 (NJEdge)	158	39	147
Other – State Government	0	2007	1947	1947	1671
Other – Local Government	0	54	54	54	
Other – Non Government	0	8	8	8	8
Total					5814

Abbreviations and Acronyms

911 Comm	New Jersey 9-1-1 Commission
IMLS	Institute of Museum and Library Services
IPEDS	Integrated Postsecondary Education Data System
NCES	National Center for Education Statistics
NJHA	New Jersey Hospital Association
NJ-DHHS	New Jersey Department of Health and Human Services

New Jersey has a strong tradition of home rule and, like many eastern states, a plethora of small governance entities – towns, townships, boroughs, cities, and other local municipalities. Among the major challenges we face in collecting broadband CAI data in the state are the dearth of strong, state-level organizations that might compel members to provide data (as opposed to comparatively weaker coordinating bodies) and the lack of existing broadband data sources. NJEdge's data on the higher education institutions to which they provide service is one of the very few such resources in the state.

NJ OIT executives worked through state-level contacts in public safety, education and libraries, etc., to encourage their constituencies to participate and submit data through the website. While some groups were more responsive than others, many expressed concerns about placing additional burdens in a time of shrinking budgets and cutbacks. Telcordia also conducted individual outreach county-by-county in the state which resulted in some additional broadband submissions from county government through the website.

We encountered a few issues with collection, interpretation and processing of CAI data:

- Some institutions provide information on multiple connections to the internet, each with its own technology of transmission and maximum speeds. These may represent separate redundant connections for a large institution that provides critical services or separate facilities for different classes of users (e.g., staff and clients). Our policy has been to submit a single entry for each institution, using the highest available download speed, but this policy may be a candidate for refinement.
- Satellite institutions such as branch libraries or campus outreach centers can complicate the CAI picture. Our policy is to attempt to collect data for each separate geographic location as a separate CAI.
- Sometimes multiple government offices are co-located in one geographic location; e.g., a large building or complex that may include county government offices, court, jail, and/or other government offices. Here the challenge is not to incorrectly overstate broadband capability or understate the need for broadband services.

- It remains challenging to convince busy employees at CAIs to take the time to provide this data.
- The CAI transfer model requires a street number and for some CAIs this is not readily available as institutions may use a cross street for directions, a PO box for paper mail, etc. We suggest that the NTIA consider making street number *optional* in the transfer model on a going forward basis.

3 Service Provider Data Reception

Telcordia defined a process for handling provider data upon receipt. The following steps describe that process:

These steps must be performed upon receipt of provider data. These steps set up the file system and database for later processing, including both the initial assessment and load, and protect the confidentiality of the information.

1. Update the provider interaction log spreadsheet with the date of receipt and other metadata.
2. Copy the email or decrypt the uploaded files to individual directory on dedicated and secure server.
3. Test that the files can be opened, read, etc. This may require using ESRI ArcCatalog to check a shapefile or file geodatabase.
4. Send an acknowledgement to the provider of receipt of readable submission, or request re-send as needed.
5. Create empty provider data report into the new folder, using the appropriate wireless or wireline template**.**
6. Connect to the PostgreSQL database and instantiate a schema for the provider
7. Import the NTIA transfer model tables to the new schema using ArcCatalog. These are available in the "ntiamodel" schema.
8. Add triggers to the newly imported tables. These triggers update columns with the user name and date/time for each insert and update.
9. Perform an initial evaluation on the submitted data, evaluating the completeness of the submission and the validity and reasonableness of the included values. Interact with provider to address any questions or issues.

4 Service Provider Data Loading

All providers are responding to the mandate to provide the different types of data that go into the various tables in the NTIA data transfer model. The provider data submissions vary in form, format and content and in the ease versus complexity of the processing and loading tasks.

In general, the most straightforward data to process are shape files submitted by wireless providers. Wireline providers who submit census block data are a step up in terms of complexity. Some cable providers simply list the municipalities which they serve. A number of smaller providers provide address lists corresponding to locations where they provide service. These are much more challenging to process as we must first manipulate the address information and then geo-code the locations; these operations can be time consuming and subject to inaccuracies.

The service provider reports attached in Appendix A give the full details per provider on all steps taken to extract, transform, and load the contents of the provider tables into the NTIA tables. Note that every NTIA table has a "shape" column where a geographic feature such as a point, line (e.g., road segment) or area (e.g., census block) must be submitted.

Here is a summary of some of our key policies and challenges:

- All non-disclosure agreements executed with providers prohibit us from disclosing customer addresses. Although some providers have not executed NDAs, we have chosen to treat all providers similarly. We have chosen to obfuscate the address data by transforming it to census blocks or street segments. This carries a slight risk of overstating coverage, but that seems more appropriate than simply dropping the data because it is sensitive.
- Speeds associated with address data from some providers represent the price plan chosen by the customer; they are definitely neither the max advertised speed nor the typical speed. Our decision was to keep the maximum speeds encountered in the census block and report them in the maximum advertised fields and to report typical as null. If customers' selections in neighboring census blocks were vastly different, we would use the highest speed in a (subjectively defined) area as the maximum advertised speed.
- Maximum advertised speed, combined with the 7-10 availability requirement, results in vagaries in interpretation. In particular, the concept of advertised speed is well suited for providers who offer services to extended areas, such as large telephone and cable television companies. Its application is less clear for smaller providers who offer

service to defined set of specific addresses. They deliver services to those specific addresses, and could offer the same service to a new tenant within the time limit. In some cases, they could increase the speed within that time period as well. They could not easily deliver service to any neighboring location with a two-week period. We have operationalized the notion of maximum advertised speed by determining the maximum speed a provider could offer on the facilities they have in place at customer locations, then reporting that speed for census blocks or street segments.

- After initial poor results in geo-coding the customer address lists provided by some cable providers who had no geo-spatial capabilities, we identified an alternate approach that leveraged the franchise-nature of cable television service in the state. We asked those cable TV providers to send us the list of municipalities that they are licensed to serve. We build the submission by locating the municipality shapes and using those shapes to find all census blocks contained within them. For large census blocks, we report all the TigerLine street segments that are contained within those blocks.
- For middle mile data, the exact definition of a connection point remains open to interpretation and requires further development. We are not completely sure that all providers interpret middle mile in the same fashion and do not have a clear enough picture ourselves to provide appropriate guidance or validation. Despite this, we have submitted the middle mile information that we received.
- All but one provider submitted 2010 Census Blocks (CBs). Xchange was the one provider who submitted 2000 CBs, requiring us to map the coverage to 2010. This results in a modest overstatement as we show availability for all 2010 CBs for which there is overlap with a 2000 CB in their serving territory.

5 Data Validation

Incoming data was subjected to a number of validation checks. When incoming data failed a validation check, we first investigated our process to ensure that we were not inadvertently creating an issue. If the problem was determined to be with the submitted data, we notified the provider concerned and recorded the interaction in the provider data report as provided in Appendix A. Where possible, we impute missing data. As reported with our April submission, we have attempted to perform some data validation using the FCC speed-test data, but had limited success due to the sparseness of the coverage of the speed-test data. Recent FCC speed test data is showing a reduction in the number of measurements, which only increases the sparsity.

We have observed a few issues that arose when processing the current submission:

- The alignment of Tiger Lines and 2010 CBs has sometimes been problematic, particularly for large CBs. When a 2010 CB has a Tiger Line road segment as part of its boundary, we have found a number of examples where there is misalignment which makes the road segment appear to be within a specific CB rather than as a boundary. Please see Section 6.1 for an example.
- New Jersey placenames can be difficult. We validate against data from the following sources: State of New Jersey geographic information (https://njgin.state.nj.us/NJ_NJGINExplorer/DataDownloads.jsp), the Federal Government placename information (http://geonames.usgs.gov/domestic/download_data.htm), and the US Postal Service data (available for a fee).
- A survey of 3100 New Jersey households was conducted in November and December by Rutgers University as Telcordia's subcontractor under this program. Householders who responded that they were broadband users were asked who their service provider was and this was compared against service provider serving areas. 95% of the responses aligned with service provider information. In the remaining 63 cases, the survey respondents reported being served by a provider whose coverage area did not appear to cover that location. Through these cases we have identified an area for additional investigation which may lead to improvements in service provider coverage. The technique, based on geo-spatial analysis of neighboring CBs is briefly described in Section 6.2.
- T-Mobile submitted wireless coverage data that provided one of the more interesting validation issues. T-Mobile provided separate information about three different varieties of 3GPP-based wireless technology, each of which supports broadband data services through mobile terrestrial wireless service capability; namely: UMTS, HSPA21 (i.e., HSPA) and HSPA42 (i.e., HSPA+)[1]. In order to avoid duplicates – that is, rows of T-Mobile data with

[1] Here are a few more technical details. UMTS is based upon 3GPP release 99 and is the oldest and slowest of the three varieties. HSPA (HSPA21) is 3GPP R6 which supports HSDPA and HSDPU for downlink and uplink high-speed packet access and offers

identical shapes and the same technology and spectrum codes, differing only in maximum speed, we performed spatial joins separately for each of UMTS, HSPA21 and HSPA42. We then submitted one shape for each 3GPP technology.

- The End_User_Category for Census Blocks or Road Segments is an optional field for designating the geography as being primarily Residential, Non-Residential, or Other (primarily neither Residential nor Non-Residential). Based on discussions with NJ OIT we have elected not to complete this field as OIT does not have a trusted data source for this information.

We applied the business rules in the script supplied by the NTIA and other data-specific validations after the data were loaded into the tables. These were applied as a check on both the data supplied by the providers and on the process we used for data collections, reception and loading.

The following business rules were applied above and beyond those in the NTIA script:

We checked uniqueness of the entries in each table, using the following definitions of uniqueness:

Layer	**Unique key**	**Notes**
Middle Mile	frn, latitude, longitude	
CAI	anchorname, address, transtech	
Census Block	frn, fullfipsid, transtech	
Street Segment	frn, tlid, transtech	Tlid is an internal column.
Wireless	frn,transtech, spectrum, shape	

We also performed the following additional validations:

Layer	**Validation Rules**
Middle Mile	• Check (dbaname, provname, frn) against our FRN reference table • Valid census block id within the state of New Jersey • Check latitude not between 38.7 and 41.4 • Check longitude not between -75.6 and -73.8 • Shape should not be empty • All check_submission rules
CAI	• Valid zip code • Check latitude not between 38.7 and 41.4 • Check longitude not between -75.6 and -73.8 • Shape should not be empty • All check_submission rules
Census Block	• Check (dbaname, provname, frn) against our FRN reference table • Valid census block id within the state of New Jersey • The area of a census block should be less than < 2 square Mile • Shape should not be empty • All check_submission rule
Street Segment	• Check (dbaname, provname, frn) against our FRN reference table • Street segment is present in a census block >= 2 square miles • Shape should not be empty • All check_submission rule

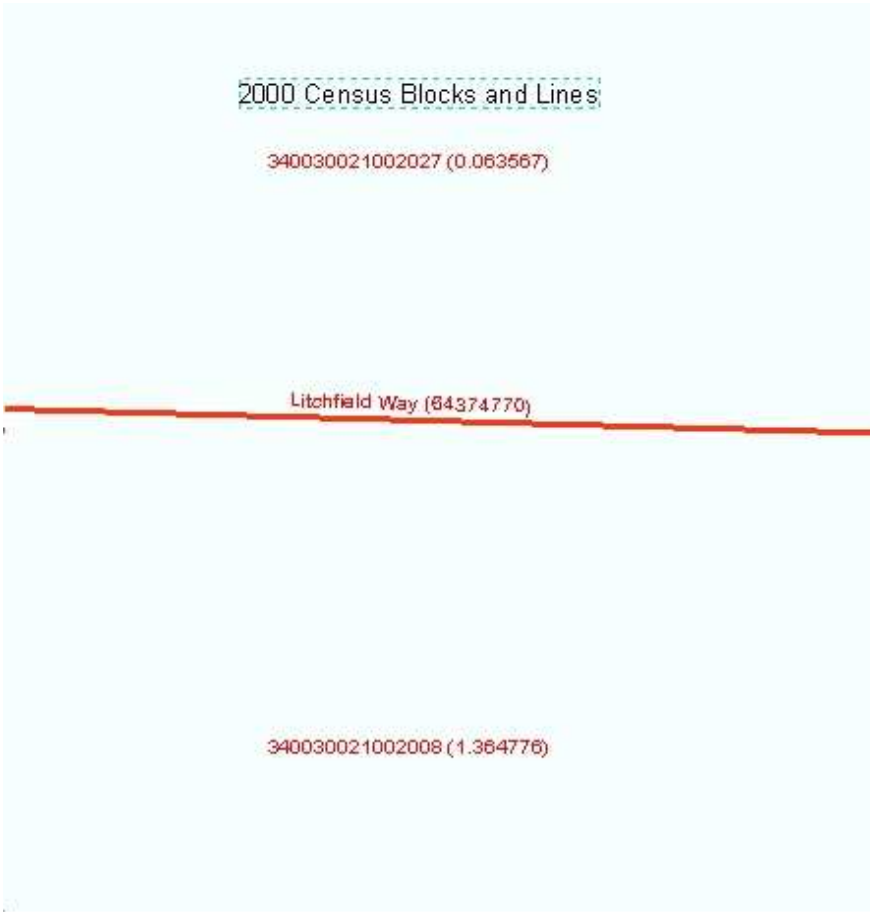
intermediate speeds. HSPA+ (HSPA42) is 3GPP R7. It is the most advanced of the three and supports high-speed packet access evolution with peak data rate increases from MIMO and higher-order modulation, among other technical advances.

Wireless	• Check (dbaname, provname, frn) against our FRN reference table • Shape should not be empty • All check_submission_rule

6 Two Issues in Geometry

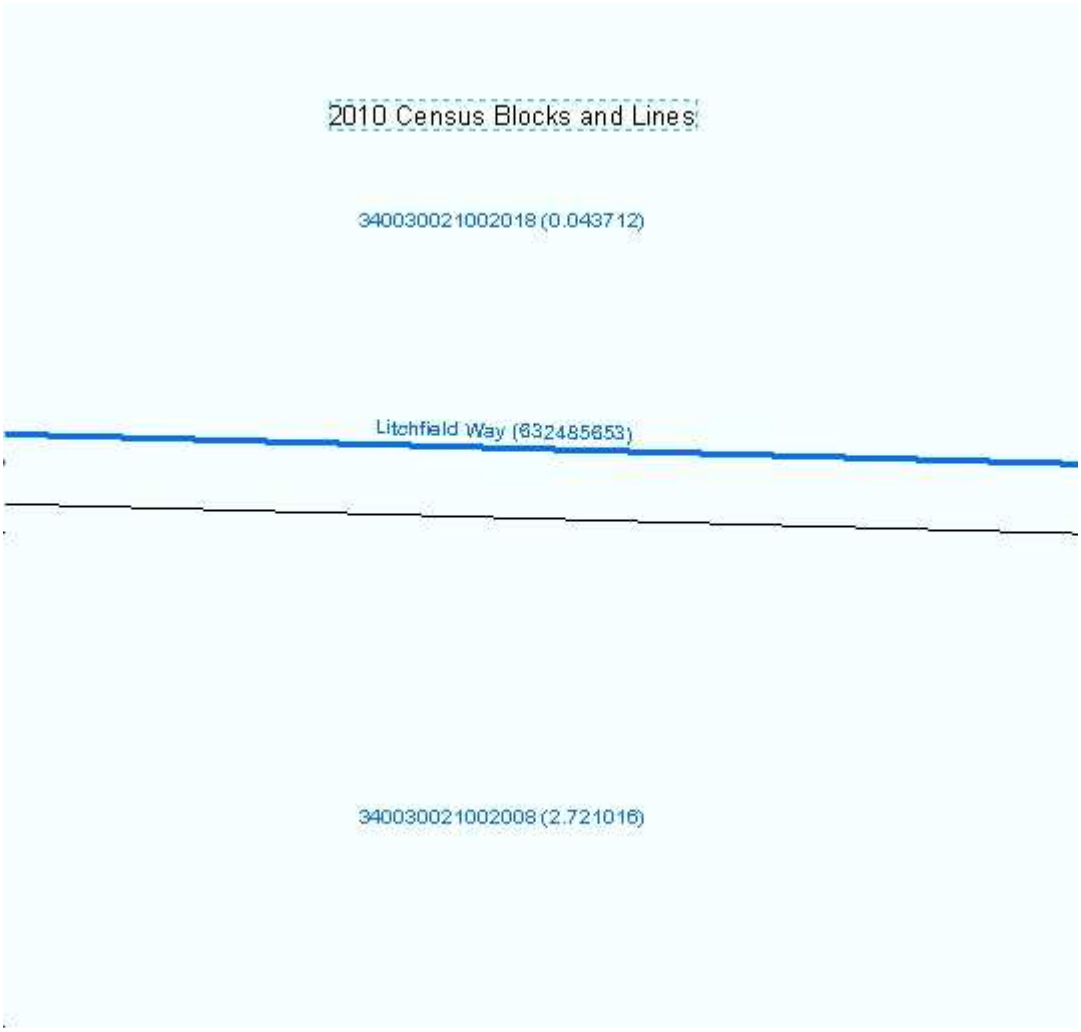
6.1 Tiger Lines and 2010 Census Block Misalignment

Here is an example of two 2000 Census Blocks and the Tiger Line which forms part of the boundary illustrating proper alignment.



The next page shows an example of the same geometry with 2010 Census Blocks and illustrates the misalignment between the line and the CB boundary.

Misaligned line and 2010 CB boundary:

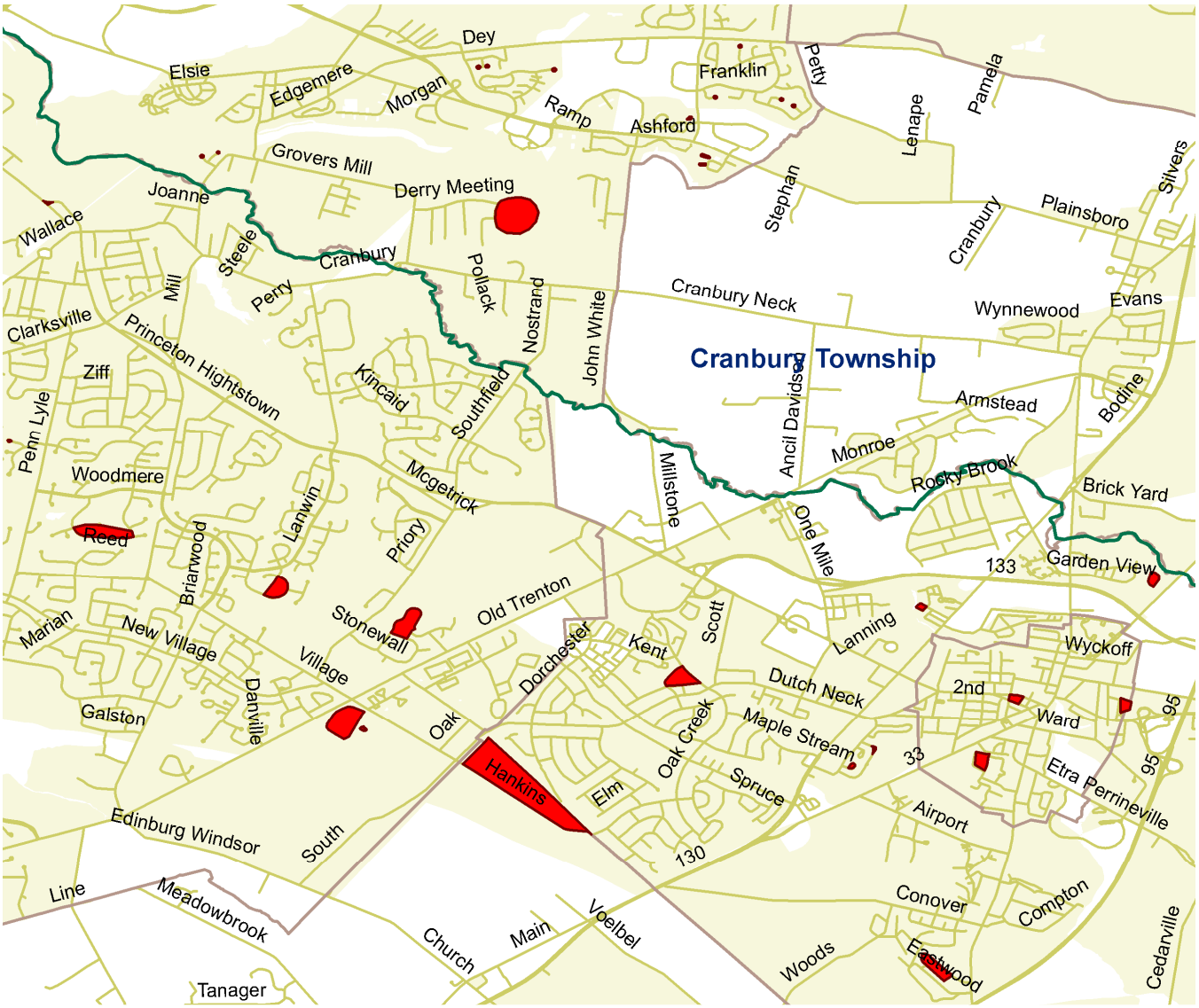


6.2 Gap Analysis of Neighboring Census Blocks

The analysis of the survey data identified some instances where a survey respondent identified their service provider and then the service provider's data did not show coverage in that respondent's Census Block. Further analysis indicated that a number of these instances occurred in 'gaps' or 'holes' in submitted provider coverage data. One way to define a simple hole is that it is a single CB that is not in the stated provider coverage area when all neighboring CBs are in the stated coverage area. Our investigations of these simple holes showed that some are associated with zero-population CBs – e.g., a CB that comprises a strip of land neighboring a major roadway. Other simple holes, however, appear to be anomalies in service

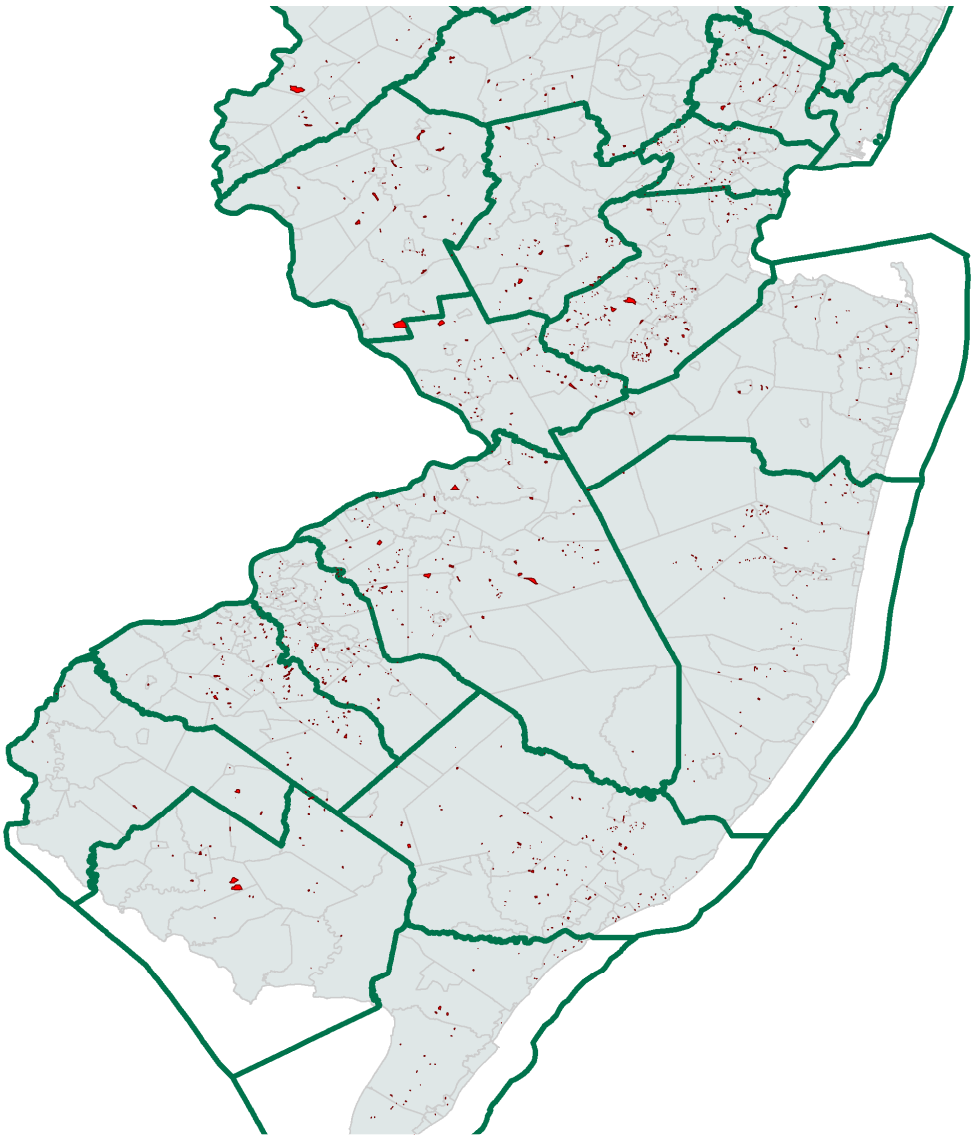
provider data as we find examples of a residential CB, surrounded by other residential CBs, and no clear rationale to explain why the initial (middle) CB would not have coverage when all neighboring CBs do have coverage.

The next figure shows a few simple holes in Comcast data from Cranbury Township at a fine resolution.



Our analysis of the simple holes shows that some are anomalies that may provide a way to improve the accuracy of provider data. To pursue such possible improvements, we developed software that automates the identification of simple holes. Somewhat to our surprise, when we ran this software on the data for this submission, we found rather sizeable numbers of holes for some of the providers. For example, we identified almost 250 simple holes for Cablevision (including Lightpath) and over 1400 for Comcast. The following graphic illustrates the simple holes for Comcast.

Graphic of Simple Holes in Comcast Data:



Given the number of holes, it is apparent that conveying them one-by-one to providers for review is not feasible. However, the identification of these simple holes opens an avenue for implementation of additional automated verification of service provider coverage. Essentially what we are considering is entering the geospatial locations of the holes in major providers' on-line service availability systems in a mechanized fashion. This would allow us to conduct an efficient and automatic internal consistency check between provider data and the web-based service availability systems offered by major providers.