New Jersey Broadband Mapping Project:

Report on Data Integration and Validation Procedures For October 2010 Submission

October 29, 2010

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 an overview of the steps taken by the New Jersey Office of Information Technology (OIT) and Telcordia Technologies to collect, receive, load, and validate broadband data submitted to us by wireless and wireline service providers. Individual provider data reports (elsewhere) provide low-level details on each provider's submission and explain how the policies presented in this document were applied to the data.

1. Provider Data Outreach

Telcordia and OIT contacted a total of 67 providers via email and telephone. OIT negotiated NDAs with those providers who required them. Once the NDAs were in place, providers were provided with instructions on data requirements and how it should be submitted via http://connectingnj.state.nj.us/.

Most providers were willing to participate, although several expressed that they felt burdened by the data collection processed asked of them. In particular, one provider declined to devote any effort to submitting data; another provider instructed Telcordia to retrieve the information from the company's web site. The large national providers clearly had processes in place to collect and submit data, while the small local providers required greater assistance. Telcordia offered assistance where possible, allowing providers to submit whatever data they had available in any convenient format. For example, some of the smaller wireline providers simply submitted a list of addresses where they offer server and some small cable operators simply submitted the names of the municipalities that the cover. While this resulted in unplanned work in the collection and processing, it enabled greater coverage of providers.

2. 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.

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 server.
3. **T**est 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.

3. Provider Data Loading

Section 1: NTIA Transfer Model Tables

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. There is a section for each of those destination tables.

The provider data submissions are generally tables, but may be text documents, Excel workbooks, or databases (shapefiles or file geodatabases). See Section 2 for special cases for treating submissions that are not tables.

These guidelines include detailed rules for extracting, transforming, and loading 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.

NTIA Table "BB_CONNECTIONPOINT_MIDDLEMILE"

This table stores points where a wireless or wireline provider has a major connection point (the exact definition of a connection point remains open to interpretation apparently). The table requires a point shape corresponding to the location, and the ID of the containing census block.

Case 1: Provider sends text data with street addresses

Translate the street addresses into (latitude, longitude) value pairs using a geocoding service. Then proceed as in Case 2.

Case 2: Provider sends (latitude, longitude) value pairs.

Create a feature class in ESRI with point shapes that correspond to the (latitude, longitude) value pairs. In that feature class, ensure the geographic coordinate system and the tolerance value are an exact match for the NTIA transfer model. Then do a spatial join against Census Bureau data to locate the containing census blocks, creating a new feature class Use that feature class to create rows with all required attributes such as Provider Name etc., and load to the target table.

Case 3: Provider sends a geodatabase with shapes.

We have not yet received shapes for middle-mile data, but if this happens, proceed as in Case 2, skipping the first step.

NTIA Table "BB_SERVICE_ADDRESS"

This table stores points where a wireline provider has a customer. The table requires a point shape corresponding to the location, and the ID of the containing census block.

We have not yet used this table. All non-disclosure agreements executed with providers prohibit us from disclosing customer addresses. Although some providers have not executed an NDA, we have chosen to treat all providers similarly. We have adopted the following policy and procedure for loading data submitted to us as customer addresses. This procedure obfuscates the address data by transforming it to census block data. This carries a slight risk of overstating coverage, but that seemed more appropriate than simply dropping the data because it is sensitive.

Uniform procedure for address level data

1. Discard PO box and other invalid addresses
2. Find (latitude, longitude) pairs by geocoding street addresses
3. Find the containing census blocks by spatial join against reference data
4. Check the size of the containing census block
    a. If <= 2 sq mi, submit block in table bb_service_censusblock
    b. Otherwise check NDA/public status of address
        i. If no NDA (i.e., public), send address (this has not yet happened)
        ii. Otherwise search for line segment.
            1. If found, submit line segment.
            2. Otherwise drop. Dropped rows go in a fallout report that is returned to provider.

As part of this procedure, duplicate records must be eliminated. Duplicates are created when multiple addresses are in the same census block. Speeds associated with address data from some providers represent the price plan chosen by the customer; they are definitely not the max advertised nor typical. Our decision was to keep the maximum speeds encountered in the census block and report them in the maximum advertised fields; report typical as null.

We will use similar logic for speed determination as defined for census blocks – look for duplicates and report highest speed among those on a street segment as the maximum advertised.

NTIA Table "BB_SERVICE_CAINSTITUTIONS"

This table stores points where a community anchor institution is located, such as a hospital, library, fire station, or police station. The table requires detailed address information, a point shape corresponding to the location, and the ID of the containing census block.

We collect this data from the institutions, not from service providers. We receive street addresses. There is no requirement to obfuscate the data. The following procedures are used to process the data:

1. Translate the street addresses into (latitude, longitude) value pairs using a geocoding service.
2. Create a feature class in ESRI with point shapes that correspond to the (latitude, longitude) value pairs.
3. In that feature class, ensure the geographic coordinate system and the tolerance value are an exact match for the NTIA transfer model.
4. Then do a spatial join against Census Bureau data to locate the containing census blocks.
5. Copy the data over.

Some institutions provide information on multiple connections to the internet, each with its own technology of transmission and maximum speeds. NTIA indicates that the primary interest is in knowing the location of CAI. Our policy is to submit a single entry for each, using the highest available download speed.

NTIA Table "BB_SERVICE_CENSUSBLOCK"

This table stores areas where a wireline provider's service is available. The table requires a polygon shape corresponding to the census block. Only census blocks that are 2 square miles or less may be reported here; no blocks larger than 2 sq mi may be loaded.

Case 1: Provider sends text data with Census Block IDs

Query the Census Bureau TigerLine reference data to locate the shape corresponding to each Census Block ID. Use that data to create rows with all required attributes such as Provider Name etc. and load to the target table.

Case 2: Provider sends a geodatabase with shapes.

Check the geographic coordinate system and tolerance value on the submitted feature class, and transform as necessary to the system used in the transfer model. Use the submitted rows to create rows with all required attributes such as Provider Name etc., and load to the target table.

NTIA Table "BB_SERVICE_OVERVIEW"

This table stores areas where a wireless or wireline provider's service is available at the county level, and includes information about subscriber weighted nominal speeds by county. The table requires a polygon shape corresponding to the county.

Some providers send the subscriber-weighted nominal speed, but others do not. The table allows nulls in most columns. Our policy for loading this table is as follow: If we have nothing to add beyond the data reported in tables bb_service_censusblock or bb_service_wireless, then do not populate the overview table. In other words, don't simply duplicate data from those other tables to populate this table. *Do* convert data if supplied at the MSA level.

Case 1: Provider sends text data with county IDs

Query the Census Bureau TigerLine reference data to locate the shape corresponding to each county ID. Use that data to create rows with all required attributes such as Provider Name etc. and load to the target table.

Case 2: Provider sends a geodatabase with shapes.

Check the geographic coordinate system and tolerance value on the submitted feature class, and transform as necessary to the system used in the transfer model. Use the submitted rows to create rows with all required attributes such as Provider Name etc., and load to the target table. (This case has not yet occurred.)

NTIA Table "BB_SERVICE_ROADSEGMENT"

This table stores streets where a wireline provider's service is available. The table requires a polyline shape corresponding to the street segment. This table should only be loaded with street segments in census blocks larger than 2 square miles.

Case 1: Provider sends text data with street names and numbers (no TLID)

Attempt to locate the road segment in a geographic feature database such as TigerLine. If found and is a good match, submit it; otherwise drop it. Dropped rows go in a fallout report that is returned to the provider.

This is an exceedingly difficult situation. Comcast submitted this type of data in June and September 2010. Matching their data to TigerLine yielded many errors. We decided to gather the list of served municipalities instead, and load data based on that.

Case 3: Provider sends text data with TigerLine segment IDs.

Query the Census Bureau TigerLine reference data to locate the shape corresponding to each segment ID. Use that data to create rows with all required attributes such as Provider Name etc. and load to the target table.

Case 3: Provider sends a geodatabase with shapes.

Check the geographic coordinate system and tolerance value on the submitted feature class, and transform as necessary to the system used in the transfer model. Use the submitted rows to create rows with all required attributes such as Provider Name etc., and load to the target table.

NTIA Table "BB_SERVICE_WIRELESS"

This table stores areas where a wireless provider's service is available. The table requires a polygon shape. No census bureau shapes or IDs are required.

The only reasonable submission for this table is a feature class in a geodatabase. We have received ESRI Shapefiles, ESRI file geodatabases, and MapInfo databases.

Check the geographic coordinate system and tolerance value on the submitted feature class, and transform as necessary to the system used in the transfer model. Use the submitted rows to create rows with all required attributes such as Provider Name etc., and load to the target table.

Section 2: Special Cases of Submitted Data

Some submissions are not compliant with the system designed by the NOFA and do not fit into any of the NTIA transfer model tables. This section discusses these cases.

Cable Providers – Municipality data: Some cable TV providers have sent 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.

Section 3: Detailed Procedures

This section presents details for manipulating the various tools to accomplish extract, transform, and load tasks mentioned earlier.

Procedure for geocoding submitted data

This procedure applies to cases in which we receive text data and must create or find corresponding shape data.

A. Submitted data was of type Lat/Long (either as a text or Excel file)
    a. shapes (points) must be manually created with ArcGIS
B. Submitted data was of type address (either as a text or Excel file)

a. addresses must be geocoded
b. shapes (points) must be manually created with ArcGIS

The following two situations require geocoding and manual ESRI tasks:

1. To obfuscate customer addresses for service availability by census block

The result of this should be a table in the geodatabase. It can be a plain table or a feature class; the tolerance value is irrelevant. Usually it's a feature class with point features that result from creating XY data based on Lat/Long value pairs, but those point features are not loaded to any bb_* table. Ideally the table has every single column submitted by the provider. The only critical column is the ID of the containing census block for the address.

If some addresses fail geocoding, the rows should be written to something suitable for returning to the provider. If some points fail the spatial join, they should be investigated for whether they are in New Jersey, whether the geocoding service made a mistake (frequently happens), or other reason.

2. To report middle mile connection points

The result of these steps should be a feature class in the geodatabase with point shapes. It must have the GCS_WGS_1984 coordinate system and tolerance value of 0.000000002. Ideally the table has every single column submitted by the provider. The critical column is the shape that corresponds to the middle mile address or lat/long, whatever was submitted, another critical column is the FIPS00, the code for the Year 2000 census block shape that contains the point. This block can be large. We filter out rows for which geocoding failed (no lat/long, no point), and also filter out rows for which the spatial join failed (such as points not in NJ).

Procedure for Loading Data (two-step import) in ESRI

The two aspects of a geographic coordinate system on a feature class are the projection and tolerance. When importing geodata from a shapefile it's easy to get the projection required by the NTIA transfer model, but impossible to set the tolerance value required by the NTIA transfer model. Further, note that ArcCatalog cannot show tolerance for a shapefile. The only way to see it is to copy the feature class out of the shapefile and into a file gdb, then look at the metadata (properties) for the feature class in the file gdb.

This procedure explains how to import data from a shapefile or file gdb called "SRC" to a feature class called "TGT" so that it exactly matches the tolerance of an existing feature class called "GOLD".

Step 1: Create a new feature class definition (i.e., import schema) in the geodatabase.

- Right click on the geodatabase, pick New, pick Feature Class
- Enter a name for the new feature class in the dialog TGT, click next
- Import a coordinate system: click import, browse to the GOLD feature class, click add, click next
- Check the tolerance value shown (it should match GOLD), click next
- Leave database storage config at default, click next
- Import the columns from feature class SRC: click import, browse to SRC, click add
- Click Finish.

Step 2: Load the data from SRC to TGT.

- Right click on TGT, pick Load, pick Load Data.
- Select the source: browse to the SRC feature class, click open, click Add, click next
- Next screen (Select the target geodatabase) accepts no changes, so just click next again.
- Verify that the field mappings are 1:1, click next again.
- Choose to load all or some of the data, click next again.
- Read the summary, click Finish

4. Data Validation

Incoming data was subjected to a number of validation checks, as outlined in this section. When incoming data failed a validation check, we notifies the provider concerned and recorded the interaction in the Provider Data Report. Where possible, we impute missing data (see for example the discussion of FRN below).

Shape

(Appears in ConnectionPoint_MIddleMile, Service_Address, Service_CensusBlock, Service_Overview, Service_RoadSegment, Service_Wireless)

Most of the tables in the NSGIC data model have a Shape column. It is the central feature of the Service_Wireless and Service_RoadSegment tables. For other tables it provides a useful but not essential addition. If Shape is provided, data on the projection used must also be provided. If the data is not supplied using the projection and tolerance values required by the NTIA transfer model, the data must be transformed appropriately.

R.1: The shape should be a polygon, segment or point (type appropriate to the table) lying at least partially in New Jersey. Portions of shapes not in New Jersey may be clipped.

R.2: Separate polygons representing the same service should not overlap.

Other consistency checks may be performed based on other data in the record.

R.3: Street segments should match the US Census Bureau 2009 Tiger Line database. Segments that cannot be matched will be reported to the provider, and may be rejected.

R.4: All projections will be converted to the system used by the NTIA transfer model. This has changed over time.

R.5: Without a projection, we have no way of obtaining the latitude and longitude of geospatial data. Geospatial data with no projection will be rejected.

Shape_Area

(Appears in Service_Census_Block, Service_Overview, Service_Wireless)

R.6: This value will be re-computed by ESRI.

Shape_Length

(Appears in Service_Census_Block, Service_Overview, Service_RoadSegment, Service_Wireless)

R.7: This value will be re-computed by ESRI.

FRN

(Appears in ConnectionPoint_MiddleMile, Service_Address, Service_CensusBlock, Service_Overview, Service_RoadSegment, Service_Wireless)

FRN is required, but may be provided as part of the name of the uploaded file or in a separate file.

R.8: If no FRN is available anywhere in the uploaded data, an FRN from the FCC database should be imputed and the provider notified to ensure that we are not attributing service incorrectly.

R.9: An explanation should be recorded when a carrier data file contains multiple FRNs. The provider need not be contacted if the explanation is clear (e.g. multiple FRNs for cable company subsidiaries.)

R.10: If the FRN is missing from individual records in carrier provided data it will be imputed, e.g. from the name of the file.

R.11: The FRN must be in the FCC FRN database [6]. Submissions with invalid FRNs will be rejected.

R.11.1: If a provider submits multiple FRN values and does not assign them to individual records, the provider should specify one of the FRN values that can reasonably be assigned to all the records.

ProvName, DBAName

(Appear in the same tables as FRN)

R.12: If ProvName or DBAName are missing, they will be set to the Business Name from the FCC FRN database.

R.12.1: The provider will be contacted for confirmation if these are inconsistent with the FCC database or the provider's Form 477 filings.

StateAbbr

(Appears in Service_Address, Service_RoadSegment)

R.13: StateAbbr should be "NJ".

R.14: StateAbbr will be imputed if missing.

R.15: Records with incorrect state abbreviations will be rejected.

StateFIPS

(Appears in Service_CensusBlock)

State FIPS codes are available from many sources, including [7].

R.16: StateFIPS should be 34 (New Jersey).

R.17: StateFIPS will be imputed if missing.

R.18: Records with incorrect State FIPS codes will be rejected.

CountyFIPS

(Appears in Service_CensusBlock, Service Overview)

New Jersey county FIPS codes are available from many sources, including [8].

R.19 CountyFIPS should be a valid county FIPS code for New Jersey.

R.20: Missing CountyFIPS codes will be imputed from other data if possible.

R.21: Records containing county FIPS codes that are invalid or missing (and cannot be imputed) will be rejected.

R.21.1: Records containing county FIPS code that do not match other data (specifically, the Census block data) will be reported to the provider. These will be still be reported to the NTIA. Providers should expect the NTIA (and by extension, Telcordia) to start rejecting such records in the near future.

BlockID, Tract, Block

(Appear in Service_CensusBlock)

The Service_CensusBlock table reports data by Census Block. The BlockID is redundant in that it can be recovered from the FIPS codes and the tract and block IDs. Definitions and documentation on census block boundaries are available from the Bureau of the Census.

The NTIA transition from 2000 Census Blocks and to 2009 Census Blocks was may be problematic for some providers. We had some providers submitting with 2000 Census Blocks and some using 2009 Census Blocks. We applied corrections to the best of our ability as follows:

R.22: If a 2000 Census Block with area greater than two square miles is split into multiple 2009 Census Blocks of less than two square miles, we will accept Service_CensusBlock entries for that block in the June 30 2010 submission.

R.23: If BlockID is missing we will impute it from the other columns, and vice versa.

R.24: Either the BlockID or StateFIPS+CountyFIPS+Tract+Block must be present. Records with neither will be rejected.

R.25: Block IDs will be checked to ensure that they are in New Jersey and are on the Census Bureau's list of valid blocks. Records with invalid blocks or blocks that are not in New Jersey will be rejected.

CBYear

(Appears in Service_CensusBlock)

As discussed above, this should be 2000 for the October 2010 submission but we will accept 2009.

R.26: Records with a value of CBYear other than 2000 or 2009 will be rejected.

R.27: If CBYear is missing it should be imputed if possible. Records where CBYear cannot be imputed will be rejected.

R.28: The value of CBYear must match the year of the Census Block in the record. Records where this is not the case will be rejected.

TransTech

(Appears in Service_Address, Service_CAInstitutions, Service_CensusBlock, Service_Overview, Service_RoadSegment, Service_Wireless)

The technology should be appropriate for the table domain. For example, in the Service_Address table this should be a code for a wireline technology. The technology should also be consistent with speed columns. For example, we would not expect to see upstream speed greater than downstream speed for an asymmetric technology.

R.29: The TransTech technology code should be taken from the Technology of Transmission domain in the NSGIC schema.

R.30: Records with incorrect or inconsistent (e.g. with respect to speed) technology codes will be rejected.

R.31: Missing technology codes may be imputed. This imputation will be documented. Records with missing technology codes may be rejected if imputation is not reasonable.

MaxAdvDown, MaxAdvUp

(Appear in Service_Address, Service_CAInstitutions, Service_CensusBlock, Service_Overview, Service_RoadSegment, Service_Wireless)

TypicDown, TypicUp

(Appear in Service_Address, Service_CensusBlock, Service_RoadSegment, Service_Wireless)

Speed should be consistent with the technology. Symmetric DSL speeds should match. Most other download speeds should exceed upload speeds (although it might be possible for the speed tiers to come out equal). All DSL speeds should be less than 50 megabits per second.

R.32: When technologies with differing values of MaxAdvDown and MaxAdvUp are summarized in the Service_Overview table, a separate record must be created for each technology.

R.33: Values of MaxAdvDown and TypicDown should be taken from the Download Speed Tier domain in the NSGIC schema.

R.34: Values of MaxAdvUp and TypicUp should be taken from the Upload Speed Tier domain in the NSCIG schema.

R.35: Records with incorrect speed codes will be rejected.

R.36: Records with inconsistent speed codes (e.g. speed codes that do not match the technology) may be rejected. The provider will be contacted for clarification.

R.36.1: Missing maximum speeds will be imputed from the Service_Overview table if possible. If this leads to inconsistent values, the records may be rejected or a more appropriate speed code imputed.

City, Zip5

(Appear in Service_Address, Service_RoadSegment)

We recognize that New Jersey placenames are problematic. 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).

R.37: Records with placenames that we cannot reasonably match may be rejected. The provider will be contacted for clarification.

R.38: Records with placenames that are not in New Jersey will be rejected. The provider will be contacted for clarification.

R.39: Records with zip codes that are not in New Jersey will be rejected. The provider will be contacted for clarification.

R.40: If both Zip5 and City are present, they should match, that is, Zip5 should be a valid zip code for City. We will accept records where City is the name of a municipality and the placename for Zip5 refers to an area lying partly within that municipality. Records where Zip5 and City do not match should be reported to the provider, but will still be reported to the NTIA if they can be geo-located.

R.41: If missing, City may be imputed from Zip5 and vice versa.

R.41.1: If missing, City and Zip5 may be imputed through geolocation.

R.41.2: Records where this is not possible should be reported to the provider, but will still be reported to the NTIA.

ID

(Appears in Service_RoadSegment)

R.42: The column labeled ID in the Service_RoadSegment table will be the Tiger Line database ID if available.

Address

(Appears in Service_Address)

The Address column in the Service_Address table is intended to be a space-separated concatenation of the other address elements.

R.43: If Address is not provided we will construct it from the other address elements, if provided.

R.44: Due to the difficulty in parsing free-format addresses we will not construct the other columns from this column.

GeogUnit, GeogUnitID

(Appear in Service_Overview)

R.45: If the unit type is "CO" (county), the ID must be a valid 5 digit county FIPS code (the first two digits must be "34", the state code for New Jersey). See the discussion of County for more on county FIPS codes.

R.46: If the unit type is "MSA" the GeogUnitID must be a valid Metropolitan Statistical Area Code [13] in New Jersey.

R.47: If the unit type is "CMA" the GeogUnitID must be a valid FCC Cellular Market Area [14] in New Jersey.

R.48: Rows with missing or invalid GeogUnitIDs will be rejected.

Longitude, Latitude

(Appear in ConnectionPoint_MiddleMile, Service_CAInstitutions)

R.49: If present, these should be in the state of New Jersey.

R.50: A publicly available geocoding service will be used to fill in missing longitude and latitude values from address data.

Spectrum

(Appears in Service_Wireless)

R.51: Submissions using the original NOFA table definition are acceptable and will be converted.

R.52: Values of Spectrum should be taken from the Spectrum Used domain in the NSGIC schema.

R.53: Records with invalid values of Spectrum will be rejected.

BHCapMidM

(Appears in ConnectionPoint_MiddleMile)

R.54: Values of BHCapMidM should be taken from the Middle Mile Backhaul Capacity domain in the NSGIS schema.

R.55: Records with invalid values of BHCapMidM will be rejected.

BHTechType

(Appears in ConnectionPoint_MiddleMile)

R.56: Values of BHTechType should be taken from the Backhaul Technology Type domain in the NSGIC schema.

R.57: Records with invalid values of BHTechType will be rejected.

Ownership

(Appears in ConnectionPoint_MiddleMile)

R.58: Values of Ownership should be taken from the Owned or Leased domain in the NSGIC schema.

R.59: Records with invalid values of Ownership will be rejected.

CAICat

(Appears in Service_CAInstitutions)

R.60: Values of CAICat should be taken from the Community Anchor Institution domain in the NSGIC schema.

R.61: Records with invalid values of CAICat will be rejected.

BBService

(Appears in Service_CAInstitutions)

R.62: Values of BBService should be taken from the Yes or No domain in the NSGIC schema.

R.63: Records with values of BBService that cannot reasonably be converted to the given domain will be rejected. (e.g. we will change case or convert yes/no and 1/0)

EndUserCat

(Appears in Service_Address)

R.64: Values of EndUserCat should be taken from the End User domain in the NSGIC schema.

R.65: Records with invalid values of EndUserCat will be rejected.

ElevFeet

(Appears in ConnectionPoint_MiddleMile)

ElevFeet is supposed to be elevation above grade level, but some providers may have interpreted it as elevation above sea level. Suspicious values are not grounds for rejection.

R.66: Suspicious values of ElevFeet (less than -100 or greater than 1500) should be brought to the attention of the provider. These will be changed to zero. The provider will be notified.

ARPU

(Appears in Service_Overview)

R.67: This column is not expected, and will be ignored if present.

Last Mile Connection Point Data

Per the NOFA clarification, this table is no longer required.

R.68: Any Last Mile Connection Point Data submitted will be ignored.

We also applied a sub-set of the NTIA business rules 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 tests were applied:

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, transtech) and (address, transtech)	
Census Block	frn, fullfipsid, transtech	
Service Overview	frn, statecountyfips, transtech	
Street Segment	frn, tlid, transtech	Tlid is an internal column.
Wireless		There may be UNIQUE rule on (frn, transtech, shape), but since the type of the shape column is ST_GEOMETRY, it is not easy to define the constraint on the column.

We also performed the following additional validations:

Layer	Validation Rules
Middle Mile	• Valid census block id • Stateabbr should be 'NJ' • Shape should not be empty
CAI	• Statecode should be 'NJ' • Valid zip code • Shape should not be empty • Transtech should not be NULL
Census Block	• The area of a census block should be less than < 2 SQ Mile • Shape should not be empty
Service Overview	• Stateabbr should be 'NJ' • Compare with the census block table and wireless table for consistency. • Shape should not be empty
Street Segment	• Shape should not be empty
Wireless	• Stateabbr should be 'NJ' • Shape should not be empty

The table below is a version of the Business Rules provided by NTIA with highlighted the rows to illustrate the tests that were performed on the data prior to submission.

1. Front-end business rules are built in the SBDD model and are enforced automatically.
2. Rules for Service Address are not implemented since we do not use the table.
3. Legend

Rule is implemented
Rule is NOT implemented
There are issues implementing and/or understanding the rule
Rule is front-end. We do nothing.

Business Rule	Layer?	Front End/Back End?	Notes
Provider Name / DBA / FRN must be consistent for all records in the entire state	Middle Mile	Back-end	Implemented by a foreign key
Ownership – valid value list of only 0 or 1	Middle Mile	Front-end	
Serving facility capacity – valid value list of only 1 – 6	Middle Mile	Front-end	
Serving facility type – valid value list of only 1 – 4	Middle Mile	Front-end	
REQUIRED COMBINATION BUSINESS RULE FOR serving capacity and serving type	Middle Mile	Front-end	
Latitude – must be a positive decimal number greater than 13 and less than 72. Must have 6 decimal places populated.	Middle Mile	Front-end	The rule is dropped in the latest model. Process is to check latitude not between 38.7 and 41.4. The topology rule also would validate it.
Longitude – must be a Negative decimal number greater than -170 and less than -60 (Except for Guam). Must have 6 decimal places populated (right hand place cannot be 0)	Middle Mile	Front-end	The rule is dropped in the latest model. Process is to check longitude not between -75.6 and -73.8. The topology rule also would validate it.
Elevation – measured in ***feet***, must be a positive number between -282 and 20,320	Middle Mile	Front-end	The assumption is that number provided is in feet. The rule is dropped in the latest model.
Point (Combination of Latitude and Longitude must fall within the state awardee submitting the value	Middle Mile	Back-end	This is the topology rule that has not been executed.
Require a FIPS Block code	Middle Mile	Front-end	This does not appear to be a front-end rule
Provider Name / DBA / FRN see Middle Mile	Service Address	Back-end	Not implemented since we do not use the service address table
FRN see Middle Mile	Service Address	Back-end	This rule is not clear
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology speed combinations (see below)	Service Address	Back-end	
Latitude business rules, see Middle Mile	Service Address	Front-end	

Longitude business rules, see Middle Mile	Service Address	Front-end	
Point (Combination) see Middle Mile	Service Address	Back-end	
The point must not list a higher technology code than the highest technology code provided in the service area overview for the same provider.	Service Address	Back-end	
The point must not provide a higher technology code than the highest technology code of any provider listed in the block (if the block is < 2 sq mi).	Service Address	Back-end	
Flag the point(s) if, the block is > 2 sq mi AND this block neighbors a block < 2 sq mi w/o service availability from any provider.	Service Address	Back-end	
The Point must be in a block which contains population.	Service Address	Back-end	
FIPS Block code	Service Address	Front-end	
BBService – valid value list	CAI	Front-end	
Latitude see Middle Mile	CAI	Front-end	The rule is dropped in the latest model. We check latitude not between 38.7 and 41.4. The topology rule also would validate it.
Longitude see Middle Mile	CAI	Front-end	The rule is dropped in the latest model. Process is to check longitude not between -75.6 and -73.8. The topology rule also would validate it.
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology speed combinations (see below)	CAI	Front-end	
Point (Combination) see Middle Mile	CAI	Back-end	
The CAI must not list a higher technology code than the highest technology code provided in the service area overview for the same provider.	CAI	Back-end	
The CAI must not provide a higher technology code than the highest technology code of any provider listed in the block (if the block is < 2 sq mi).	CAI	Back-end	
Flag the CAI(s) if, the block is > 2 sq mi AND this block neighbors a block < 2 sq mi w/o service availability from any provider.	CAI	Back-end	

If CAI must be in a block with other service.	CAI	Back-end	
The CAI must be in a block that contains population	CAI	Back-end	
Provider Name / DBA / FRN see Middle Mile	Census Block	Back-end	Implemented by a foreign key
FRN see Middle Mile	Census Block	Front-end	This rule is not clear. It is defined as back-end in other places
StateFIPS – valid value only	Census Block	Back-end	
CountyFIPS – valid value list only	Census Block	Back-end	
Tract	Census Block	Back-end	
Block	Census Block	Back-end	
CBYear – Valid value list is 2000 only	Census Block	Back-end	This column has been dropped
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology speed combinations (see below)	Census Block	Front-end	
Combination StateFIPS, CountyFIPS, Tract, BlockGroup, Block need to be in the acceptable range of blocks < 2 sq mi (we need to produce this list) for that state	Census Block	Back-end	
The block must touch (e.g. be a neighbor with) at least one other block < 2 sq mile with availability	Census Block	Back-end	
Is the dissolve of the Block data for speed the same as the service overview for speed for that provider?	Census Block	Back-end	
The block must contain population	Census Block	Back-end	
Provider Name / DBA / FRN see Middle Mile	Service Overview	Back-end	Implemented by a foreign key
FRN see Middle Mile	Service Overview	Front-end	This rule is not clear It is defined as back-end in other places
GeoUnit – valid value list only (County only)	Service Overview	Back-end	Pre-fill to county. I think it is a front-end rule
GeogUnitID – valid value list only	Service Overview	Back-end	Change to StateCountyFIPS (5 characters)
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology speed combinations (see below)	Service Overview	Front-end	
Speed business rule to check nominal weighted speed?	Service Overview	N/A	

Provider Name / DBA / FRN see Middle Mile	Street Segment	Back-end	Implemented by a foreign key
FRN see Middle Mile	Street Segment	Back-end	This rule is not clear
AddMin must be less than AddMax	Street Segment		
StateAbbrev – valid value list	Street Segment	Front-end	Check if stateabbr = 'NJ'
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology speed combinations (see below)	Street Segment	Front-end	
The data must be in a block > 2 sq mi?	Street Segment	Back-end	
Flag the data, if the data is in a block that does NOT neighbor a block < 2 sq mi with service (e.g. all neighbor blocks that are < 2 sq mi have no availability)?	Street Segment	Back-end	
Flag the data if there is no neighbor with block with availability?	Street Segment	Back-end	
Provider Name / DBA / FRN see Middle Mile	Wireless	Back-end	Implemented with a foreign key
FRN see Middle Mile	Wireless	Back-end	This rule is not clear.
TransTech – valid value list only	Wireless	Front-end	
REQUIRED COMBINATION BUSINESS RULE FOR transmission technology / spectrum / speed combinations (see below)	Wireless	Front-end	

5. Overview, Challenges and Mitigation

Overview

Sixty-five service providers representing 57 active, independent companies were contacted to determine eligibility for this program. Thirty companies were identified as facilities-based providers, meeting the 10 day service provision window. Of these 30 in-scope companies:

- 28 provided data included in the October 2010 submission.
- One provider submitted data too late for this submission.
- One provider declined to participate due to lack of internal resources to gather and process the data.

Provider Name	NDA?	Supplied Data	Data Verified and Submitted?
Advanza Telecom Inc	not required	yes	yes
AT&T Mobility LLC	yes	yes	yes
Broadview Network Holdings	yes	LATE	no
Cavalier Telephone Mid-Atlantic LLC	yes	yes	yes
CenturyTel, Inc. (CenturyLink)	yes	yes	yes
Cogent Communications Inc.	not required	yes	yes
Comcast Cable Communications, LLC	yes	yes	yes
CSC Holdings (Cablevision)	yes	yes	yes
DIECA Communications (Covad)	yes	yes	yes
Global Online Electronic Services		yes	yes
Hometown Online	not required	yes	yes
Hotwire Communications		Refused	No
Hughes Network Systems	not required	yes	no (satellite)
Leap Wireless (also Cricket)	not required	yes	yes
Monmouth Telephone & Telegraph	yes	yes	yes
Netlogic (Voxitas)	yes	yes	yes
One Communications Corp	yes	yes	yes
RCN NY Communications	yes	yes	yes
Service Electric Cable, Hunterdon	not required	yes	yes
Service Electric Cable, Sparta	not required	yes	yes
Sprint Nextel	yes	yes	yes
StarBand Communications	not required	yes	no (satellite)
Time Warner Cable	yes	yes	yes
T-Mobile	yes	yes (T)	yes
tw telecom holdings	not required	yes	yes
Verizon	yes	yes	yes
Verizon Wireless	yes	yes	yes
Wave2Wave Communications	yes	yes	yes
WildBlue Communications	yes	yes (satellite)	no (satellite)
XO Communications	yes	yes	yes

Primary Challenges and Mitigations

Our major challenges and mitigation strategies are as follows:

- Address geocoding, particularly in census blocks greater than 2 square miles, has been challenging due to the quality of provider data and timeliness of available street map information. To mitigate this we purchased the latest ESRI street database and have provided direct feedback to providers on problematic addresses. For cable providers, we also made use of NJ BPU data and identified census blocks in municipalities.
- The time required for outreach and technical interaction with service providers proved substantially more extensive than anticipated. Numerous outreach attempts by email and phone were needed at many companies to secure cooperation and obtain data. The data arrived in very diverse formats and structures, due not only to differences in company operations, but also to differing interpretations of the NOFA requirements. This created two challenges – first, significant technical interchange was required to obtain additional information and correctly interpret the data. Secondly, we had to develop customized data processing procedures for each submission. Our mitigation strategy was simply to shift our effort, emphasis, and resources to this activity over the summer months.
- Community Anchor Institution (CAI) identification and outreach was time-consuming and response rates to date are low. Our strategy going forward is to work through state-level organizations such as the NJEDge and the State Library Association. We will be implementing a web-based data submission capability on our website (http://connectingnj.state.nj.us/).