Virginia Broadband vendors

Broadband vendor data was delivered to the CIT and VGIN in different time intervals and in different formats for the fall 2010 Broadband data reporting. Those formats received were GIS Shapefiles, Excel Spreadsheets, Text Files, and numerous other images and resources to highlight input data.

Data sources and database environment

Virginia manages and updates a statewide GIS Road Centerline data layer sourced from local jurisdictional 911 data, with output products released to state, local, and commercial participants on a quarterly basis. This data provides end users access to information more current than most commercial and federal datasets. To leverage the state road centerlines, The Commonwealth of Virginia has also been collecting address point data from localities to increase geocoding accuracy. Both of these GIS data sets from Q32010 were imported into a SQL server GIS database with unique ids to relate associated non-spatial vendor data. Virginia Road Centerlines and Address Points both have a unique ID which was used in output joins. Both data sets reside in a Virginia based projection and coordinate system and were converted to WGS84 for the broadband mapping project.

Two Geocoding services were created and used for outputting spatial data to points; one using the VA road centerlines and one using VA Address points. Features that needed to be geocoded were done separately to ensure accuracy and maintain consistency.

Per NTIA request, The 2000 geography for Census Blocks were downloaded through TIGER and a generic integer ID (VGIN block ID 2000) was given to each census block in the GIS data in order to join address level data, although more importantly for record management. The Full FIPS ID generated by the USCB was text and although it was used in joining block data if provided by a vendor, it was not used in joining other features. The following are GIS data sets used and the Unique ID used or created for production in the SQL server database:

Virginia RCL GIS layer = V_LEID
Virginia Address Point GIS layer = AP_ID
Census Block GIS layer = VGIN_BLOCK_ID_2000

Before Spatial data was loaded into Broadband SDE database, it was imperative for Virginia Address points to have a reference to block information. As a pre-processing step, The Virginia address points were spatially joined to the Census block layer and the VGIN_BLOCK_ID_2000, as well as the census block id number added to the individual address point. This was used in selecting by feature per block.

Related provider data was set up to be stored in non spatial SQL server tables and included the feature id in order to create a relationship to the GIS feature class.
A related provider table was created in SQL server for each spatial feature type that was available for use in this project. NTIA provided File Geodatabase for data modeling and for schema production. Therefore all addressing or census block data was used in the format provided in the model.

ESRI Spatial views were created as production outputs and field names were updated in SQL server to match the NTIA data model schema. This allowed flexibility in the naming conventions that were already put into place, i.e., the source data did not need to be renamed. Of the spatial views, the address point, street centerline, and census block field names were renamed in SQL server for each spatial view in order to match schema information requested in the NTIA output loading data model.

Vendor processing

A total of 36 vendors participated in the VA Broadband mapping effort for the Fall 2010 submittal. Several had no updates since the summer report (Virginia's first dataset was delivered to NTIA in June 2010) and were recalculated based on the previous submission. Vendors with changes reported new data sets; No vendors reported deltas only.

Each Vendor was given a location on the VGIN server under F:\VGIN\GIS_Development\Broadband\2010_Vendor_Data_Fall with the vendor name as its containing folder. Inside of the vendor folder, a processing personal Geodatabase was created in order to manage the data provided and to convert all input and output data to a legible GIS format. The Geodatabase used a naming convention such as "VENDOR NAME_2010_Staging" in order to keep organization and standardization of data entry for each vendor.

There were several types of data reported to CIT and VGIN that needed processing in order to be aggregated for loading into the NTIA reporting Geodatabase:

- 13 vendors provided only addresses
- 12 vendors provided both census blocks and road segment ranges
- 6 vendors provided only wireless information
- 3 vendors provided only census blocks
- 2 non-wireless vendors provided GIS data

All vendors participating provided speed information for census block, road centerline segment, or addresses. Depending on what data type was provided, VGIN followed several processing methods:

Only address number was provided

Address number data was provided in tabular form and almost all of the vendors that provided address level data had the same vendor number, DBA Name, FRN, Transmission Technology, and Upload/Download for all of their customers in the database provided. Sizes of customer databases were very similar between vendors who provided only address number information and no census block information. The individual vendor spreadsheet was imported into the vendor's staging database as an address spreadsheet. The vendor data was then geocoded to both the Address Point locator and Road Centerline locator and two GIS outputs were created; one point feature class which were geocoding results of a custom address point locator and one point feature class which were geocoding results of a road centerline address point locator.

Both address and road centerline points were placed on the appropriate feature. In order to track the Virginia segment or address information though time, the house or address range segment unique feature ID was needed. A spatial join was performed in order to capture the feature identifier. This output feature class was either exported or calculated and exported into a feature class in the vendor's staging Geodatabase and named "AP_SDE_Load" or "RCL_SDE_Load" depending on the feature geocoded result and selection process. The specific tables were loaded into the SDE master non-spatial tables for Address Points and Road Centerlines and records were then added to each spatial view for loading into the NTIA model.

Census blocks less than two square miles and address ranges were provided

Generally a spreadsheet provided for census blocks less than two square miles from a vendor had the census block ID along with additional necessary attributes for reporting (Provider name, DBA name, FRN, Transmission, download and upload speeds) for blocks less than two square miles. This data was joined to the Census Block GIS layer in ArcMap and only join matched results were selected to be returned. The generated VA block ID was captured through this methodology and blocks were exported to the vendor's staging Geodatabase and named "blocksless2". These newly created blocks were then exported to a non-spatial table called "CENSUS_SDE_Load" in the vendor's staging Geodatabase and then loaded into SDE to the block non-spatial table in order to be added to the census block spatial view.

For segments greater than two square miles where address ranges were provided, block ID was provided by all vendors per segment. This block FIPS id was joined to the statewide census blocks SDE table in order to create a polygon of blocks greater than two square miles. The polygon was exported to a feature class in the vendor's staging Geodatabase and called "blocksmore2". Address points which joined to the blocks by block ID were selected to get an address point sample for the blocks greater than two square miles. Also, road centerlines were clipped by the "blocksmore2" block feature class to get potential segments to use in selections and block information was calculated to the segments output. These points and centerlines were exported to the vendor's staging Geodatabase with the purpose of creating subsets of the entire feature classes.

Depending on the number of blocks greater than 2 square miles provided by the vendor and the amount of features selected compared to the original report, some vendor points and centerlines extracted were exported to “AP_SDE_LOAD” or “RCL_SDE_LOAD” and used in reporting.

If the vendor reported ZIP code information as well as road name in their range data for blocks greater than two square miles, records were selected which satisfied the query of Road name in the VA Road Centerlines like the attributes of the vendor and zip equal to the attributes of the vendor. The results were well matched and exported to SDE_RCL_LOAD and loaded into the SDE master non-spatial table.

After several selection methods were attempted to whittle down results for Block ID and range, one method seemed to work. If ranges did have valid information for street name and address range, a query of the points and centerlines created was performed based on the provided vendor spreadsheet to get a GIS representation of the potential coverage. Results in address points that satisfied the query of block full FIPS ID (provided in vendor data) equal to the provided attribute and road name (provided in vendor data) like the provided attribute were then selected. These resulting points would then be geocoded to the lines through the RCL geocoding service and lines would be selected based on their location to get linear features for point inputs. The results were to be output to AP_SDE_LOAD in the vendor’s staging Geodatabase and loaded into the SDE master non-spatial table for Spatial View & loading. There were several localities in Virginia who did not provide points so the centerlines needed to be clipped based on block area and centerlines needed block FIPS id placed on the segment to perform a selection. The segments which satisfied a query for block full FIPS ID (provided in vendor data) equal to the provided attribute and road name (provided in vendor data) like the provided attribute were to be output to “RCL_SDE_LOAD” in the vendor’s staging Geodatabase and loaded into the SDE master non-spatial table.

Wireless Coverage

Wireless providers generally submitted data to CIT in shapefile form. If a vendor used a grid system to display area of interest and additional redundant polygons were present, the vendor’s speed data was dissolved to get a single polygon per provider or at least get the fewest amount of polygons. Some wireless providers only sent a customer database and in the case of those vendors, they were treated the same as those who submitted address only data.

Only census blocks less than two square miles were provided

Generally a spreadsheet provided from a vendor had the census block ID along with additional necessary attributes for NITA reporting (Provider name, DBA name, FRN, Transmission, download and upload speeds). This data was joined to the Census Block GIS layer and in ArcMap and only join match results were returned. The generated VA block ID was captured through this methodology.

The results were selected and the selection was output to its own feature class in the vendor's staging Geodatabase and called "blocksless2". If the attributes needed adjustment or fields needed to be renamed to match the data model, they were done in the output feature class. The attributes of this spatial output were exported again to a non spatial table in the individual vendor's staging Geodatabase and named using the naming convention "CENSUS_SDE_Load". This table was then loaded into the SDE master census table and records were then added to the spatial view for loading into the NTIA data model.

If GIS data was provided by a non-wireless vendor

The bulk of vendors who provided GIS data were wireless providers.
Several vendors that were not wireless carriers did provide GIS coverage data via GIS polygons for census blocks and lines for road segments along with their reporting excel or text files. The polygon data provided was represented by blocks less than two miles, therefore block ID was captured and features were spatially joined to the master census feature class and outputs were converted to "CENSUS_SDE_Load" in the vendor's staging database. The newly created table was loaded to the non-spatial SDE table for output spatial view.

Linear features provided by vendors who reported GIS data represented segments in blocks greater than two square miles. If a vendor provided linear GIS data from a commercial data source or the TIGER lines, the Virginia RCL was used in order to select segments which matched road centerline coverage. Select by location was used in order to extract VA road centerlines within 5 meters of the provided data. This was to provide actual 911 road name as well as local geometry based on the most recent Virginia RCL. Attributes for broadband were then conflated to the selections based on location.

Both census blocks less than two square miles and address numbers provided

No vendors reported information in this format for the Fall 2010 release. This is the preferable format of submission for future releases due to the minimal effort of processing needed.

Additional vendor data sets used in reporting data

Community Anchor Institution information

Virginia Tech held speed tests in order to get download and upload speeds for Community Anchor institutions. NTIA requested that the data model not be changed so unfortunately speed data was not reported since VT had the typical upload and download speeds populated. The requested attributes were advertised upload and download speeds; the vendors who provided service to the Anchor institutions in Virginia were unreported in many categories therefore advertised speeds were unavailable to report.

Middle Mile

If middle mile data was provided by the participating vendor, this was converted into a Geodatabase table in the providing vendor's staging Geodatabase. All appropriate fields were added in order to load data spatially. This information was exported to an access database and then exported to comma delimited text for the non spatial middle mile reporting feature class.

Pricing

If nominal weighted subscriber speed was available from a vendor, the data was placed into an Access database which followed the format of requested text output information from NTIA. It was then output to a requested tab delimited text file for the release.

Speed based on CMA/MSA/RSA

If speed was available by cellular market area or MSA/RSA and provided to CIT and VGIN, this information was placed into an Access database which followed the format of requested text output information from NTIA. It was then output to a requested tab delimited text file for the release. This category of reporting was least provided by vendors.

Issues/Considerations

Cellular Market Area reporting

There were several cases where Cellular Market Area (CMA) coverage was used in reporting speed tiers. A CMA shapefile was located online and loaded into SDE for processing. The features used had CMA id number and CMA name and this information was spatial joined to blocks, streets, and addresses where pertinent.

Several major providers submitted a census blocks less than two square miles and road segment ranges although left the upload and download speeds out. The information for maximum advertised for up and download speed was provided in a separate spreadsheet and referenced speeds by cellular market area (CMA). A CMA shapefile was downloaded and used potential census block information and this contained reference information to other spreadsheets. The advertised download and upload speeds were provided by cellular market area. The road centerline information was then spatially joined to this updated table as well as the address points to get the speed information.

Data version issues & future processing

Several vendors reported blocks in 2008 geography so block FIPS id did not join from the provided block spreadsheet to the statewide block file. Since these changes reported no results when joining, the column where block was reported was formatted to allow only

the exact amount of characters for the block FIPS id to be equal from the provider to the feature selected. This seemed to be the cause of why some blocks were reported in the less than two square mile feature class but their geometry was actually more than two from 2000.

Vendors who reported address level data only did not have any additions or subtractions done to their reported data. It was geocoded directly and results were loaded into the master data set wholesale so one thing that may be done or at least need to be looked at for the next submission is where these vendors report addresses that fall in blocks less than two square miles. There may be many cases where these vendors actually need a block placed in the report instead of centerlines and points. This may reduce the amount of total address points and road segments submitted as well as increase the individual coverage area for a vendor if we do leave out features that sit on top of these polygons.

With further analysis of road name and address range data provided by vendors, many roads had null values for address range high and low as well as the street name. This was problematic when attempting to select by attributes since only data yielded results for segments where this information was populated.

Additional findings showed that several vendors provided TIGER TLID in their reporting for census blocks larger than two square miles. This was found after the fall release. For future releases, it may yield more accurate results for reporting purposes to use the TIGER ID joined to the vendor database where TLID is available for segments where census blocks are greater than 2 square miles. Creating output centerline data sets to use with selection by location instead of selection by attributes would be ideal in working with two separate geometry sources (VA centerlines vs. TIGER lines) and will be considered when viewing the data for the next release. In that regard, lines would match with the TIGER lines and the information could then be conflated to the Virginia Road Centerlines by geographic identification.

Also, further analysis of the data may show which centerline source vendors are using. Geocoding data may be a step that is only needed for address data only in the future and not needed for ranges if the census data can be utilized for a geographic feature.

Records Management

An excel spreadsheet checklist was created to keep track of what vendors were participating in the Fall 2010 mapping project as well as a format each vendor provided its data in and what method was needed in order to get the data into a GIS format. This spreadsheet kept record of the Vendor Name, the type of data each vendor provided, the type of output of GIS data that occurred from the data provided, If the vendor provided middle mile data, If the vendor provided pricing data, if the vendor provided pricing information, if the vendor provided data by CMA, and if the GIS data was

exported to non spatial data and loaded into the master vendor table for addresses and RCL relationship.

Another field in the management spreadsheet was created to note whether the coverage was either Actual or potential. Actual coverage was noted if the vendor provided address level data in a spreadsheet which translated to the entire current coverage area of the vendor. Potential coverage was noted if the vendor provided census blocks or GIS data to show that their coverage may extend beyond certain geographic extents.

The NTIA came out with a datapackage.xls which performed the same tasks as our records management checklist. Moving forward with additional submissions, it may be better to use the NTIA document to ensure data is kept up to their needs.