Data Processing Methods

Provider Participation

In Round 5, the California Public Utilities Commission identified 225 potential broadband providers, 150 of whom did not submit data, and 75 who did. These providers comprise over 99.9% of the total broadband connections in California, according to data contained in the latest FCC Form 477.

Data Collection

The California Public Utilities Commission (CPUC) sent out a Data Request to broadband providers to initiate the Round 5 data collection. Potential providers were strongly encouraged to submit broadband service availability data. Providers who previously submitted data were also sent maps displaying their Round 4 coverage and validation results to guide their 5th round submissions. Data submission instructions were posted online to assist providers along with template files, sample shape files and record formats on the CPUC Broadband Mapping Website at:

http://www.cpuc.ca.gov/PUC/Telco/Information+for+providing+service/BroadBand+Mapping.htm

The data submission instructions point each provider to the wireless and/or wireline datasets, which are separated into sections for those with GIS data (shape files or file geodatabases) and those without GIS data (text or Excel files). For providers with GIS capabilities, statewide census block and TIGER/Line shape files were provided on the CPUC website. The square mileage of each block was calculated in advance in the sample census block shape file. Using the shape files, providers were able to determine which blocks in their footprint were less than two square miles and which were two square miles or greater and therefore needed to be represented using the road segment shape file. For providers without GIS capabilities, Excel spreadsheets were provided incorporating record field formats adhering to the NOFA data submission requirements.

Community Anchor Institutions (CAI)

CAI data initially came from the eligible entries of California Teleconnect Fund (CTF) program. The CTF program provides a 50% discount on telecommunications bills for qualifying schools, libraries, government-owned and operated hospitals and health clinics, and other community based organizations. The CAI addresses were geocoded to point locations and loaded into a file geodatabase. Technology of transmission and speed data were included and identified either through information received from the institutions themselves (as in the case of libraries), or from those service providers who responded to our request for such information. To provide CAI ID information (as in the case of schools), we used the California Department of Education search engine website (http://www.cde.ca.gov/re/sd/).

CPUC Initial Data Verification

Each data set submitted by broadband providers was reviewed against the GIS data model posted on the SBDD Network website, and checked if mandatory fields were filled in, and if each field contained the appropriate range of values. Where possible, we made certain that appropriate field headers were used and that each field contained the correct data type. When data was found to be missing or incorrect, the provider was contacted and the issue was documented in a separate provider spreadsheet. Some providers submitted high quality data sets, while others gave incomplete, unexpected, or incorrect data. New information, correspondence with the providers, and fixes made by the CPUC were also documented in each provider spreadsheet.

Chico GIC Geoprocessing

After the initial CPUC review, data was transferred to the Geographical Information Center (GIC) at CSU Chico for geocoding, geomatching, propagation of wireless service by antenna, and validation of geographic data. In those cases where the CPUC received street address level data from broadband providers, such addresses were assigned a point location, (geocoded) and then geomatched to census blocks and street segments.

AT&T submitted 5th Round speed information at the county level as they have done in previous rounds. As in the 4th Round, their 5th round data was adjusted in two ways by the Chico GIC using the most recently available Form 477 data and AT&T wire center locations from a purchased database. First, in census tracts where AT&T's U-Verse service is enabled, AT&T's raw Form 477 data was used to extract the highest speed tier reported to have customers in the census tract containing each block/road segment AT&T serves. The 477 speed tier was used as the max advertised downstream speed in the block/road segment instead of the tier actually submitted. Second, to determine what speed tiers should actually be used in non-U-Verse enabled tracts, AT&T wire center locations from a purchased database were used to employ a degradation model based on loop length to estimate the highest speed ADSL service that could be achieved based on distance from the wire center.

Wireless providers who were unable to submit a shape file or geographic representation of their service area provided tabular system, tower, and antenna information. Wireless parameters were used to model the service area, and from that we created a shape file. The wireless propagation model is based on the Longley-Rice, Irregular Terrain propagation model. Individual unit specifications are used to measure performance based on frequency, transmit power, receiver sensitivity, antenna gain, and height. Signal coverage patterns are produced for each individual unit taking into account terrain and vegetation features that may hinder signal dispersion.

CPUC Final Data Verification

The resulting datasets were delivered from Chico to the CPUC in the SBDD transfer model geodatabase for final review and verification. Data sets were checked again and reviewed for unexpected changes resulting from the geocoding /geomatching process. Geoprocessed data was visually reviewed using

ArcGIS to verify service area footprints, and the SBDD check submission Python script was run on each dataset to identify unexpected values.

Deliverable Data

The final dataset is delivered to the NTIA/FCC in file geodatabase format with the following feature classes:

- BB_ConnectionPoint_LastMile – not required per Clarification to the NOFA.
- BB_ConnectionPoint_MiddleMile – Point between the local "last mile" network and the middle mile network which goes on to connect to the internet backbone. This is a confidential dataset.
- BB_Service_Address – not included per the CPUC NDA.
- BB_Service_CAInstitutions – Community Anchor Institutions: points geocoded from address lists
- BB_Service_CensusBlock – Broadband availability polygons for areas less than 2 square miles
- BB_Service_Overview – Service overview by County including Subscriber Weighted Nominal Speed
- BB_Service_RoadSegment – Broadband availability line segments for areas 2 square miles and greater
- BB_Service_Wireless – Wireless service area polygons.

Planned Validation Methods

The following validation methods will be conducted on Round 5 data. Detailed maps showing submitted service area footprints and areas that could not be validated will be distributed to each provider for feedback.

FCC Form 477

FCC Form 477 collects information about broadband connections to end user locations, wired and wireless local telephone services, and interconnected Voice over Internet Protocol (VoIP) services, in individual states at the Census Tract level. A shape file was created for each provider reflecting the availability of broadband service at each census tract where the provider reported customers of their fixed broadband service. These layers were used to cross reference ISP data submissions to the CPUC.

ID Insight, BroadBand Scout

BroadBand Scout is a third party, comprehensive and unbiased dataset specifically designed to show the carriers, connectivity, speed and usage details of the national broadband landscape. ID Insight's patent-pending process analyzes hundreds of millions of internet transactions that link a consumer's physical address to their internet carrier. BroadBand Scout data is provided as tabular point locations geomatched to the census block level less the two square miles in area and to the street segment level where census blocks are greater than two square miles in area. A shape file was created for each provider reflecting the presumed availability of broadband service at each census block or street segment where Broadband Scout reported online customer transactions. These layers were used to cross reference ISP data submissions to the CPUC.

TeleAtlas Wire Center and Wire Center Region

The Wire Center Premium product is a comprehensive database for mapping and analyzing wire center service areas. It forms the backbone of the Tele Atlas® Telecommunication Products line. This product lists every Local Exchange Carrier (LEC) landline wire center in the United States. The term "wire center" refers to the location where the telephone company terminates the local lines; this is usually the same location as a central office, although a wire center might house one or more central offices. Buffers were created at 12,000 feet and 18,000 feet from provided Wire Center point datasets to cross reference ISP data submissions to the CPUC. The wire center boundary is a representation of the area served by all of the switching equipment housed at that physical location. Wire Center Region polygon GIS layers were provided and used for cross referencing ISP data submissions to the CPUC.

FCC Consumer Broadband Test (Non-Mobile App)

The FCC Online Consumer Broadband Test collects information regarding the location of the client, the engine used to provide the speed test, download speed, upload speed, latency, jitter, packet loss, minimum round trip time, maximum round trip time, and average round trip time at a specified point location. A shape file was created to represent each location at which speed tests were performed based on geocoded address records. All point locations were then geomatched to the census block level where less the two square miles in area and to street segment level where census blocks are greater than two square miles in area. These layers were used to cross reference ISP data submissions to the CPUC where sub-broadband speeds were reported and/or where there were no tests performed.

FCC Consumer Broadband Test (Mobile App)

The FCC Mobile Consumer Broadband Test collects information regarding the location of the client, the client's operating system, the engine used to provide the speed test (always OOKLA for mobile tests), download speed, upload speed, and latency, at a specified point location. A shape file was created to represent each location at which speed tests were performed based on latitude and longitude coordinate pairs. All point locations were then geoprocessed to the census block level where less the two square miles in area and to street segment level where census blocks are greater than two square miles in area. These layers were used to cross reference ISP data submissions to the CPUC where sub-broadband speeds were reported and/or where there were no tests performed.

FCC Broadband Dead Zone Reporting Form

The FCC offers a Broadband Dead Zone Reporting Form for recording any address or city level queries done using the National Broadband Map that either failed to return any providers at the specified location, or is a location which a user knows has no service. The FCC Broadband Dead Zone Form collects information regarding the location of the client, whether the client has internet access at their home, what type of internet access the client has at their home, and whether or not the client would be interested in purchasing broadband internet if service options were available. A shape file was created to represent each location for which dead zone forms were filled out based on geocoded address records. All point locations were then geomatched to the census block level, where less than two square miles in area, and to street segment level, where census blocks are greater than two square miles in area. These layers were then used to cross reference ISP data submissions to the CPUC where dead zones and/or no services provided were reported.

California State Map Broadband Service Survey Feedback

The CPUC offers the Broadband Service Survey within its interactive map. The survey records user feedback based on address, city, or zip code level queries against the State's Broadband Availability. It collects information regarding the location of the client, whether the client is accessing the internet from their home, place of business, or any other location, whether or not the client purchases broadband service, and if not, why they choose not to purchase broadband service. A shape file based on geocoded address records was created to represent each location for which service surveys were submitted where the respondent indicated non-subscription because of no broadband availability. All recorded locations were then geomatched to the census block level, where less than two square miles in area, and to the street segment level, where census blocks are greater than two square miles in area. These layers were then used to cross reference ISP data submissions to the CPUC

Chico GIC Data Validation Processes

Each individual provider's data was validated independently using all applicable validation methods. The following fields were added to each individual provider's data tables as follows to record validation results and to allow symbology of discrepancies based on validation methods for further interaction with each provider to refine their data submissions.

- FCC_477 (FCC Form 477)
- BBSCOUT (ID Insight BroadBand Scout)
- TA_WC_REG (TeleAtlas Wire Center Region)
- WC_VAL_12K (TeleAtlas Wire Center 12,000 foot buffer)
- WC_VAL_18K (TeleAtlas Wire Center 18,000 foot buffer)
- VAL12k_18k (TeleAtlas Wire Center 12,000 to 18,000 foot buffer ring)
- DEGRAD_FT (TeleAtlas Wire Center distance)
- FCC_TST (FCC Consumer Broadband Test Non-Mobile App)
- FCC_MOBL (FCC Consumer Broadband Test Mobile App)
- FCC_DZ (FCC Broadband Dead Zone Reporting Form), and
- CA_SRVY (State Map Broadband Service Survey Feedback)

The final step was a summary statistics report of all validation results for all submitted providers. Summary statistics include validity counts and percentages for all validation methods, specific to provider and technology.

Wireline Census Block and Street Segment Validation

A spatial selection was performed on Census Block and Street Segment data, either submitted by the provider, or created from submitted address records through a geocoding/spatial selection process, to derive only those blocks or street segments which intersect polygons in a given validation layer. Counts are recorded as number of unique blocks or unique segments which share geographic area with any given validation layer, compared to the total number of unique blocks submitted by, or created for, a given provider. Percentages are recorded as percentage of the total number of unique blocks or street

segments which share geographic area with any given validation layer, compared to the total number of unique blocks submitted by, or created for, a given provider.

Wireless Validation

A spatial selection was performed on Wireless Availability data, either submitted by the provider, or created from tower and antenna location information, to select only those polygons which intersect a given validation layer. Results are recorded as a percentage of the total geographic area of wireless coverage sharing geographic area with any given validation layer compared to the total coverage area submitted by, or created for, a given provider.