

# California Broadband Data Processing and Validation

Data as of December 31, 2018

---

**Contents**

- Overview ..... 3
  - Definitions ..... 3
- Data Collection ..... 3
- Initial Data Verification ..... 4
- Geo-processing ..... 4
- Data Validation ..... 5
  - Validation Data Sources ..... 5
  - Validation Terms ..... 5
  - Validation Methods ..... 6
    - General ..... 6
    - Subscription by Census Block ..... 6
    - Public Feedback ..... 8
- Validation Results ..... 10
  - Highest Category ..... 10
  - Red-zone/Purple-zone ..... 10
- Post Validation Data Verification ..... 11

## Overview

This document summarizes how the broadband data collected by the CPUC during 2019 was processed and validated. For the December 31, 2018 data collection cycle, we received broadband deployment and subscription data from *[inset final count]* providers. This represents over 99.9% of the total reported residential broadband connections in California (based on a comparison with all data collected by the FCC).

## Definitions

The CPUC uses the FCC broadband definitions described below:

**Broadband:** Lines (or wireless channels) that terminate at an end-user location and enable the end user to receive information from and/or send information to the Internet at information-transfer rates exceeding 200 kbps in at least one direction.

**Broadband Deployment:** Fixed broadband connections are deployed in a Census Block if the provider does, or could, within a service interval that is typical for that type of connection—that is, without an extraordinary commitment of resources—provision two-way data transmission to and from the Internet with advertised speeds exceeding 200 kbps in at least one direction to end-user premises in the Census Block. Companies that would rely on the ordering or installation of a not-yet-leased circuit (including unbundled network elements defined in 47 C.F.R. § 51.319, TDM-based connections, or packet-based connections) to provide service in a Census Block not currently served, should NOT treat that Census Block as having service. Dark fiber acquired under an Indefeasible Right of Use (IRU) should be considered the “owned” facilities of the company that acquired the IRU when the dark fiber is used as part of that entity’s own system.

## Data Collection

Pursuant to Public Utilities Commission Decision [D.16-12-025](#), all communications providers certificated and/or registered with the California Public Utilities Commission (CPUC), that also file Form 477 with the Federal Communications Commission, shall submit annually to the Communications Division by April 1st, broadband subscriber and deployment data at a *Census Block level* as of the prior calendar year’s end in a form as designated by Communications Division Staff. Only Mobile providers may submit broadband subscriber data at the Census Tract level. All providers may fulfill the subscriber reporting requirement by submitting subscriber data at the more granular street address level.

Broadband data is to be submitted in the formats posted on the [Broadband Mapping Program](#) website. Please download the data formats, workbooks, and/or shapefiles appropriate for your submission.

### Broadband Data Submissions

All Fixed and Mobile Broadband Deployment data submissions are submitted in accordance with the file formats provided on the CPUC’s [Guidelines for Broadband Data Submission](#) webpage.

## Initial Data Verification

Each data set collected from broadband providers is reviewed against the GIS data templates posted on the CPUC website and checked if mandatory fields are filled in, and if each field contained the appropriate range of values. Where possible, we made certain that appropriate field headers are used and that each field contained the correct data type. When data is found to be missing or incorrect, the provider is contacted. These concerns include tech type mismatch, errant or missing block codes, lack of headers, missing speeds and other mistakes that require correction for the data to be validated. Below is a list of rules each submission must conform to before geoprocessing begins.

### Data Verification Parameters

- Each census block code must be a valid 15 digit California census block code.
- The DBA Name must be uniform across all records.
- The FRN code must be 10 digits and uniform across all records.
- The technology code must be 10, 11, 12, 20, 30, 40, 41, 42, 43, 50 or 70.
- If the Consumer field = 1 the MaxAdDn and MaxAdUp fields must be non-zero.
- If the Business field = 1 the CIRdn and CIRup fields must be non-zero.
- For subscription by address, full street addresses must be provided. No PO Boxes, descriptive locations or missed house numbers. Example (*123 Main Street, Anytown, CA 98765*).

Additionally, submitted downstream and upstream speeds are checked against an expected theoretical maximum speed specific to the technology specified. If a submission contains speeds that exceed this maximum, the provider is contacted for explanation.

## Geo-processing

After the initial data integrity review by CPUC staff, the data is transferred to the Geographical Information Center (GIC) at CSU Chico for geo-coding, geo-matching, and validation of geographic data. In cases where the CPUC received street address level data from broadband providers, such addresses were geocoded and aggregated to census blocks.

### Coastal Buffer Census Blocks

Records submitted for any water-only, unpopulated census blocks along the California Coastline are removed.

### Consumer vs Business Indication

Provider deployment census blocks for consumer and business are split into separate outputs. Only consumer (residential) level deployment is validated and utilized for CPUC's determinations for eligibility for the California Advanced Services Fund (CASF) grant program.

In deployment records where the consumer field indicator value is '1', the record is added to the Consumer Output. In deployment records where the business field indicator value is '1', the record is added to the Business Output. The result creates two separate outputs with the fields listed to the right.

Consumer Output	Business Output
BlockCode	BlockCode
DBA	DBA
FRN	FRN
TechCode	TechCode
Consumer	Business
MaxAdDn	CIRdn
MaxAdUp	CIRup

Mobile providers were asked to submit shapefiles depicting coverage. These coverage areas were used to limit the extent of the mobile interpolations based on statewide mobile broadband testing. The results from the interpolation model were then substituted for the provider submitted coverage shapefile. For mobile providers not tested under the statewide mobile testing, the submitted coverage shapefile was used.

After each provider’s data was geo-processed, a shapefile of the coverage and speeds was uploaded to a unique online account with provider-specific login credentials. The CPUC shared the unique login credentials with each provider so they could view their geo-processed data prior to validation.

## Data Validation

The CPUC validates broadband deployment data against broadband subscription and public feedback. Broadband deployment is removed in census blocks where the presence of service is unable to validate. For census blocks where we were unable to validate advertised speeds, we indicate it as an optional “purple zone” on our interactive map’s representation of individual provider service areas. Areas we were unable to validate do not mean there is no service, or that service at a particular speed is not available, rather, it means that we were unable to confirm the presence of service based on the data sources available to us. The next section lists the various data sources we use to validate broadband deployment and speeds.

## Validation Data Sources

The table below summarizes the validation method, data type, and to which type of broadband connection the validation method applies.

Data Source	Data Type	Fixed	Mobile
Broadband Subscription by Census Block	Number of subscribers by upstream and downstream speeds by census block used to validate availability and speed at census block. Broadband subscription provided at the address level is aggregated to census blocks.	YES	NO
Public Feedback	Public feedback reports by address. Address data is aggregated to census blocks. Used to validate lack of service. Reports of “no service” result in removal of that census block from provider’s coverage.	YES	NO

## Validation Terms

**Validation Resource:** Source material used in comparison with broadband deployment data to verify the existence of coverage and advertised speeds (i.e. Subscription at Census Block).

**Speed Pass:** Validation resource has spatial overlap with provider submission; Validation resource speeds meet or exceed provider submitted maximum advertised downstream and upstream speeds.

**Coverage Pass:** Validation resource has spatial overlap with provider submission, but Validation resource speeds are below provider submitted maximum advertised downstream and upstream speeds.

**No Coverage:** Validation resource has no spatial overlap with provider submission or Validation resource indicates no coverage exists.

**No Validation Resource:** No validation resource for provider.

## Validation Methods

### General

- Subscription validation is processed per provider; the validation data is used only when the provider name and technology in the validation data corresponds to the name and technology of the provider.
- Subscribed technology codes of 10, 11 and 12 are combined into the Tech Name “ADSL” for the purposes of validation and can validate any deployed technology codes of 10, 11, and 12 (i.e. a subscription to code 12 can validate the deployment of code 11).
- Subscribed technology codes of 40, 41, 42 and 43 are combined into the Tech Name “CABLE” for the purposes of validation and can validate any deployed technology codes of 40, 41, 42 and 43 (i.e. a subscription to code 42 can validate the deployment of code 41).
- Public Feedback validation is processed per provider, but is not technology specific.
- Public Feedback validation results take precedence over subscription data.
- Subscription data submitted as address locations are geocoded, aggregated to census blocks and processed using the Subscription by Census Block workflow.

### Subscription by Census Block

#### Subscription by Block field = Subscr\_blk

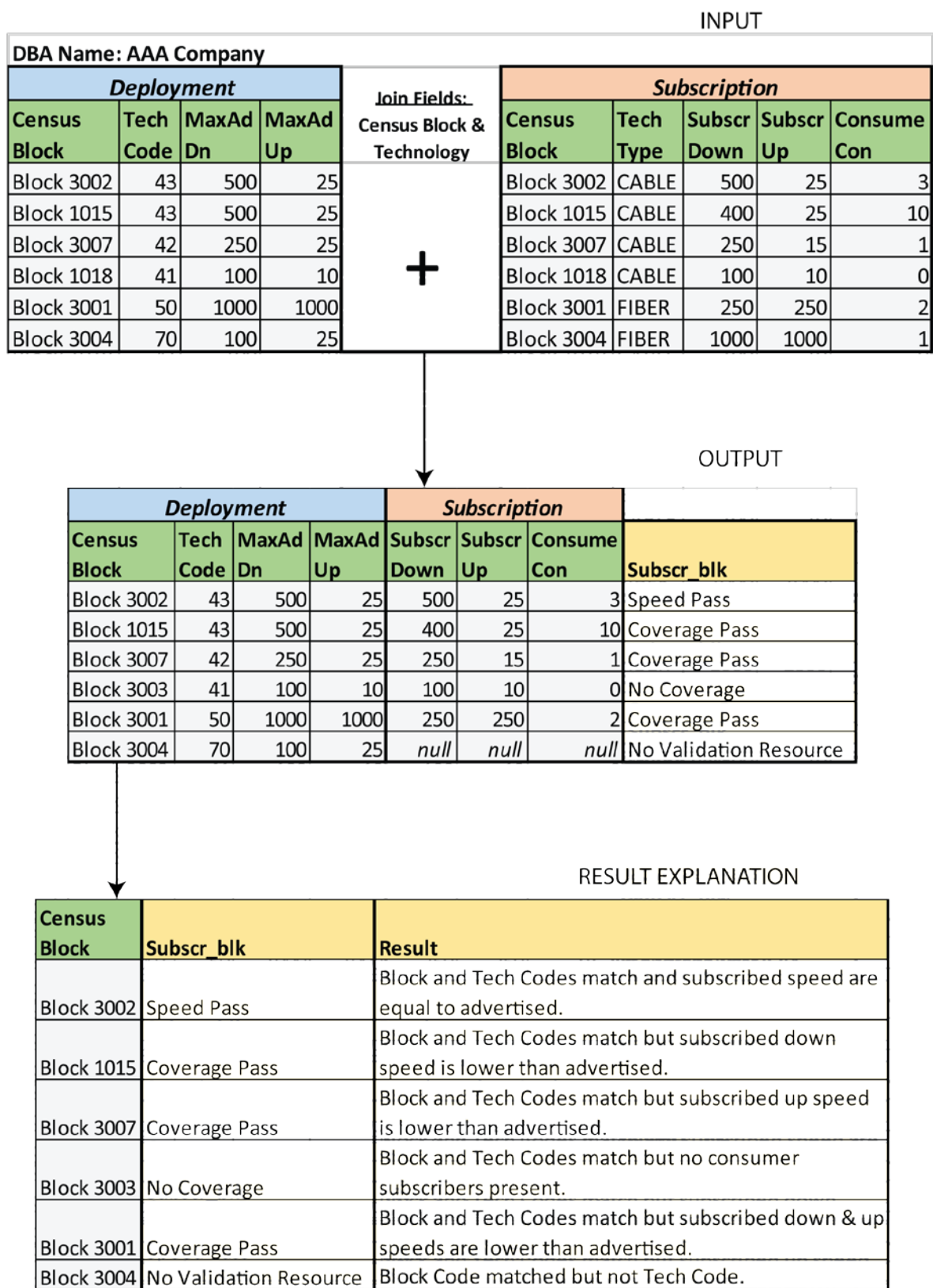
Subscription data is processed for all fixed broadband providers with any of the following technologies: Asymmetric DSL, ADSL2/ADSL2+, VDSL, Symmetric DSL, Other Copper Wireline, Cable Modem – Other, Cable Modem DOCSIS 1/1.1/2.0, Cable Modem – DOCSIS 3.0, Cable Modem DOCSIS 3.1, Optical Carrier/Fiber to the End User, and Terrestrial Fixed Wireless.

Subscription by block is submitted by the provider. Subscription data submitted at the address level was geocoded and aggregated to the census block level. Only consumer (residential) subscriptions are used to validate consumer deployment, unless determined that there is no technical difference between business and consumer service. For data as of Dec. 31<sup>st</sup> 2018, the CPUC determined that all Fixed Wireless (Technology Code 70), both consumer and business connections could be used for the purpose of validating consumer deployment. For each ISP, deployment data is joined to subscription by census block code and validation results are then coded into the Subscr\_blk field. This process is completed separately for each unique technology. Results are coded as follows:

1. **Subscr\_blk = 2 (Speed Pass):** The downstream and upstream subscribed speeds are equal to or greater than the highest downstream and upstream deployed speed in the corresponding block.
2. **Subscr\_blk = 1 (Coverage Pass):** The downstream and/or upstream subscribed speeds are less than the highest downstream and upstream deployed speed in the corresponding block.
3. **Subscr\_blk = 0 (No Coverage):** No subscribers in deployed corresponding block.
4. **Subscr\_blk = -1 (No Validation Resource):** No subscription data is available for the provider or technology listed in deployment.

Figures 1 illustrates the subscription by census block process for fixed deployment submissions.

Figure 1.



## Public Feedback

### Public Feedback field = Pubic\_Fdbk

Public Survey data is processed for all fixed broadband providers with any of the following technologies: Asymmetric DSL , ADSL2/ADSL2+, VDSL, Symmetric DSL, Other Copper Wireline, Cable Modem – Other, Cable Modem DOCSIS 1/1.1/2.0, Cable Modem – DOCSIS 3.0, Cable Modem DOCSIS 3.1, Optical Carrier/Fiber to the End User, and Terrestrial Fixed Wireless.

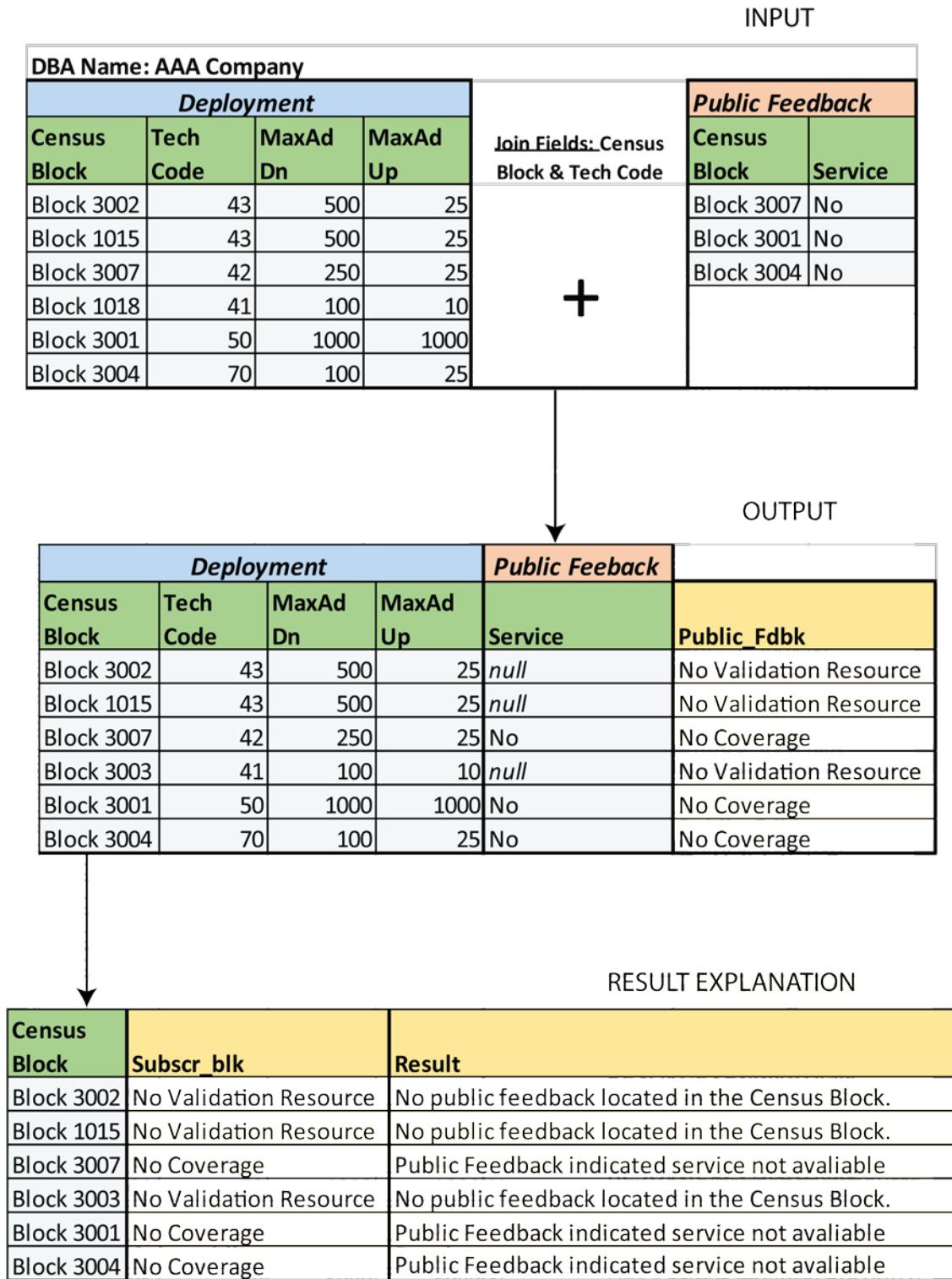
Public feedback from the CPUC Interactive Broadband Map and paper forms are geocoded and aggregated to the census block level. Only public feedback indicating denials of service within a three year period are used for the purposes of validation. For data as of December 31<sup>st</sup> 2018, public feedback can be no older than December 31<sup>st</sup> 2015. The public feedback data at the census block level is joined to deployment data and validation results are then coded into the Public\_Fdbk field. Results are coded as follows:

1. **Pubic\_Fdbk = 0 (No Coverage):** Public Feedback in corresponding census block indicated service was not available by the ISP.
2. **Pubic\_Fdbk = -1 (No Validation Resource):** No Public Feedback has occurred within the last 3 years which indicates service not available.

Figure 2 illustrates the Public Feedback validation process for fixed deployment submissions.



Figure 2.



## Validation Results

### Highest Category

#### Highest Category field = HIGH\_CAT

Highest Category is populated with a combination of the validation outputs from the Subscr\_blk and Public\_Fdbk fields. This is done using the following methodology.

- If Public\_Fdbk = 0 then HIGH\_CAT = 0.
- If Public\_Fdbk = -1 then HIGH\_CAT = Subscr\_blk.

The highest category field (“HIGH\_CAT”) represents the final results of validation.

### Red-zone/Purple-zone

#### Red-zone

Records where the Highest Category field was coded as “0” (No Coverage) are “Red-zones”. A Red-zone is an area for which no subscriber census block could validate the providers claim to service the area, or a public feedback report indicates service in that block is not available. The deployment census block will be removed from the provider’s final coverage.

#### Purple-zone

Records where the Highest Category field was coded as “1” (Coverage Pass) are “Purple-zones”. A Purple-zone is an area for which no subscriber census block could validate the providers claimed service speeds but service has been validated at a lower speed.

Figure 3 illustrates the calculation of the Highest Category and the Red-zone/Purple-zone assignment.

Figure 3.

Deployment							
Census Block	Tech Code	MaxAd Dn	MaxAd Up	Subscr_blk	Public_Fdbk	HIGH_CAT	Red/Purple-zone
Block 3002	43	500	25	Speed Pass	No Validation Resource	Speed Pass	
Block 1015	43	500	25	Coverage Pass	No Validation Resource	Coverage Pass	Purple-zone
Block 3007	42	250	25	Coverage Pass	No Coverage	No Coverage	Red-zone
Block 3003	41	100	10	No Coverage	No Validation Resource	No Coverage	Red-zone
Block 3001	50	1000	1000	Coverage Pass	No Coverage	No Coverage	Red-zone
Block 3004	70	100	25	No Validation Resource	No Coverage	No Coverage	Red-zone

## **Post Validation Data Verification**

Upon completion of the initial geo-processing and validation, broadband data submissions are loaded into a private web viewer for a provided to each provider to review and provide feedback should they believe that any areas were included or excluded in error and the assigned analyst may also have recommendations for changes to their Broadband Data Submission to more accurately reflect their coverage. Data is then revised based on recommendations for changes from the analyst before being considered final. Final data is then used to population the California Interactive Broadband Map.