California Broadband Data Processing and Validation

Data as of December 31, 2020

Contents

Overview

This document summarizes how the broadband data is collected, processed, and validated by the California Public Utilities Commission (CPUC), Broadband Mapping Program. For the December 31, 2020, data collection cycle, 175 providers were contacted to provide current broadband deployment and subscription information. New data submissions were received from 141 providers. An additional 23 providers who were non-responsive to the data request had their data from their December 31, 2019, submission reused. Reuse of provider data was only done as a last resort and where the commission felt it necessary. A total of 164 providers are included in the December 31, 2020, data collection cycle.

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.

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. For the current data collection cycle, the Commission elected to collect Satellite broadband deployment and subscription data limited to new service being deployed by Space Exploration Technologies Corp. (SpaceEx), using the fixed broadband data formats.

Initial Data Verification

For this data collection cycle, the CPUC implemented the use of an online data collection portal. Each broadband ISP was provided login credentials to submit their data. Data was automatically review for data integrity 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. Errors were automatically flagged for the ISP prior to their dataset being uploaded and they were prompted to correct. 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. Mobile providers were asked to submit shapefiles depicting coverage. No automatic data verification was performed on mobile providers.

Data Verification Parameters for Fixed and Satellite Providers

- 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, 44, 50, 60 or 70.
- If the Consumer field = 1 the MaxAdDn and MaxAdUp 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 by the data submission portal, 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. Any additional issues or irregularities found in the data at this time are reported back to CPUC staff to contact the ISP directly.

Coastal Buffer Census Blocks

Records submitted for any water-only, unpopulated census blocks along the California Coastline are removed for the purposes of display on the California Interactive Broadband Map, or other CPUC map.

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

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

Table 1.

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
Broadband Subscription by Census Tract	Number of subscribers by upstream and downstream speeds by census tract used to validate availability and speed of corresponding mobile coverage area.	NO	YES
CalSPEED Mobile crowdsourced results	Speed test data from CalSPEED mobile app, recorded within 6 months of 12/31/2020, limited to LTE capable devices and WiFi results removed. Speed test data is aggregated to census blocks and used to validate availability and speed.	NO	YES

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.

Fixed 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, 12 and 20 are combined into the Tech Name "DSL" for the purposes of validation and can validate any deployed technology codes of 10, 11, 12 and 20 (i.e. a subscription to code 12 can validate the deployment of code 11).
- Subscribed technology codes of 40, 41, 42, 43 and 44 are combined into the Tech Name "CABLE" for the purposes of validation and can validate any deployed technology codes of 40, 41, 42, 43 and 44 (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 in address format are geocoded and 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, Cable Modem DOCSIS 4.0, 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, 2020, the CPUC determined that all Fixed Wireless (Technology Code 70), both consumer and business connections would 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.

INPUT

DBA Name: AAA Company

Deployment				Join Fields: Census Block & Technology	*Subscription*				
Census Block	Tech Code	MaxAd Dn	MaxAd Up		Census Block	Tech Type	Subscr Down	Subscr Up	Consume Con
Block 3002	43	500	25		Block 3002	CABLE	500	25	3
Block 1015	43	500	25		Block 1015	CABLE	400	25	10
Block 3007	42	250	25		Block 3007	CABLE	250	15	1
Block 1018	41	100	10	+	Block 1018	CABLE	100	10	0
Block 3001	50	1000	1000		Block 3001	FIBER	250	250	2
Block 3004	70	100	25		Block 3004	FIBER	1000	1000	1

OUTPUT

Deployment				*Subscription*			
Census Block	Tech Code	MaxAd Dn	MaxAd Up	Subscr Down	Subscr Up	Consume Con	Subscr_blk
Block 3002	43	500	25	500	25	3	Speed Pass
Block 1015	43	500	25	400	25	10	Coverage Pass
Block 3007	42	250	25	250	15	1	Coverage Pass
Block 3003	41	100	10	100	10	0	No Coverage
Block 3001	50	1000	1000	250	250	2	Coverage Pass
Block 3004	70	100	25	*null*	*null*	*null*	No Validation Resource

RESULT EXPLANATION

Census Block	Subscr_blk	Result
Block 3002	Speed Pass	Block and Tech Codes match and subscribed speed are equal to advertised.
Block 1015	Coverage Pass	Block and Tech Codes match but subscribed down speed is lower than advertised.
Block 3007	Coverage Pass	Block and Tech Codes match but subscribed up speed is lower than advertised.
Block 3003	No Coverage	Block and Tech Codes match but no consumer subscribers present.
Block 3001	Coverage Pass	Block and Tech Codes match but subscribed down & up speeds are lower than advertised.
Block 3004	No Validation Resource	Block Code matched but not Tech Code.

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, Cable Modem DOCSIS 4.0, 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, 2020, public feedback can be no older than December 31, 2017. 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.

INPUT

DBA Name: AAA Company

Deployment				Join Fields: Census Block & Tech Code	*Public Feedback*	
Census Block	**Tech Code**	**MaxAd Dn**	**MaxAd Up**		**Census Block**	**Service**
Block 3002	43	500	25		Block 3007	No
Block 1015	43	500	25		Block 3001	No
Block 3007	42	250	25	+	Block 3004	No
Block 1018	41	100	10			
Block 3001	50	1000	1000			
Block 3004	70	100	25			

OUTPUT

Deployment				*Public Feeback*	
Census Block	**Tech Code**	**MaxAd Dn**	**MaxAd Up**	**Service**	**Public_Fdbk**
Block 3002	43	500	25	*null*	No Validation Resource
Block 1015	43	500	25	*null*	No Validation Resource
Block 3007	42	250	25	No	No Coverage
Block 3003	41	100	10	*null*	No Validation Resource
Block 3001	50	1000	1000	No	No Coverage
Block 3004	70	100	25	No	No Coverage

RESULT EXPLANATION

Census Block	**Subscr_blk**	**Result**
Block 3002	No Validation Resource	No public feedback located in the Census Block.
Block 1015	No Validation Resource	No public feedback located in the Census Block.
Block 3007	No Coverage	Public Feedback indicated service not avaliable
Block 3003	No Validation Resource	No public feedback located in the Census Block.
Block 3001	No Coverage	Public Feedback indicated service not avaliable
Block 3004	No Coverage	Public Feedback indicated service not avaliable

Fixed 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 overall final result 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

Mobile Validation Methods

General

- Subscription validation is processed per provider; the validation data is used only when the provider name in the validation data corresponds to the name of the provider.
- Mobile subscription is used to validate all mobile technologies and spectrums equally.
- CalSPEED test locations are aggregated to the census block level. The average measured speed per provider is used to validate the deployed speed.
- Since mobile deployment data is not represented by census geography, validation data is not joined by a census ID. Instead, validation materials are spatially intersected with the mobile deployment.

Subscription by Census Tract

Subscription by Tract field = Subscr_tra
Subscription data is processed for all mobile broadband providers.

Subscription by tract is submitted by the provider. For each ISP, deployment data is spatially intersected with subscription by census tract and validation results are then coded into the Subscr_tra field.

1. **Subscr_tra = 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 tract.
2. **Subscr_tra = 1 (Coverage Pass):** The downstream and/or upstream subscribed speeds are less than the highest downstream and upstream deployed speed in the corresponding tract.
3. **Subscr_tra = 0 (No Coverage):** No subscribers in deployed corresponding tract.
4. **Subscr_tra = -1 (No Validation Resource):** No subscription data is available for the provider.

CalSPEED Mobile Crowdsource Results

CalSPEED results field = CalSPEED
CalSPEED data is processed for all mobile broadband providers.

CalSPEED data is crowdsourced from the CalSPEED mobile application for Android and iOS. CalSPEED results are aggregated to census blocks per ISP. For each ISP, deployment data is spatially intersected with CalSPEED results by census block and validation results are then coded into the CalSPEED field.

1. **CalSPEED = 2 (Speed Pass):** The downstream and upstream speeds measured by CalSPEED are equal to or greater than the highest downstream and upstream deployed speed in the corresponding block.
2. **CalSPEED = 1 (Coverage Pass):** The downstream and/or upstream speeds measured by CalSPEED are less than the highest downstream and upstream deployed speed in the corresponding block.
3. **CalSPEED = 0 (No Coverage):** No CalSPEED results in corresponding deployment.
4. **CalSPEED = -1 (No Validation Resource):** No CalSPEED data is available for the provider.

Mobile Validation Results

Highest Category

Highest Category field = HIGH_CAT
Highest Category is populated with a combination of the validation outputs from the Subscr_tra and CalSPEED fields. This is done using the following methodology.

- MAX value between Subscr_tra and CalSPEED = HIGH_CAT.
    - Example. Subscr_tra = 2 and CalSPEED = 1 then HIGH_CAT = 2

The highest category field ("HIGH_CAT") represents the final result 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.

Post Validation Data Verification

Upon completion of the initial geo-processing and validation, broadband data submissions are loaded into a private web viewer. Each provider is given unique login credentials to review and provide feedback on their resulting coverage 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 reflect their coverage more accurately. 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, determine eligibility for the CASF program, generate broadband adoption statistics, and perform analysis on broadband availability to Californians.

Table 2 contains a breakdown of the final provider tallies compared with the previous year data collection.

Table 2.

Technology	Number of Providers for Dec. 2020 Submission (new)	Number of Providers for Dec. 2019 Submission (old)
Fixed	158	157
Mobile	5	5
Satellite	1	0
Total	**164**	**162**