2022

FCC Broadband Serviceable Location Fabric

METHODS MANUAL

COSTQUEST ASSOCIATES, INC.

Authored By:

CostQuest Associates
1430 E. McMillan St.
Cincinnati, OH 45206
U. S.

Publication Date:

October 2022

Fabric Version 1

Table of Contents

The Broadband Serviceable Location Fabric (BSLF or Fabric) is a geospatial dataset of all structures in the United States which have or may require fixed broadband service.

Created under the direction of the Federal Communications Commission (FCC), the Fabric is a special-purpose geospatial dataset supporting the Broadband DATA Act (BDA). [1] The purpose of this manual is to describe the mechanics of a twice-yearly Fabric development process. Equally important is a discussion of the context behind design choices.

This manual consists of four sections. The first section describes the background and context around Fabric development. It includes a discussion of the critical community of users who will contribute to, correct, and use the Fabric. The second section provides an overview of the Fabric data model and terminology. The third section describes the mechanics of Fabric[2] production. The fourth section describes the output files provided to users.

1. Background and Context

The Broadband Serviceable Location Fabric is a special-purpose dataset designed to meet the needs of a specific use case.

1.1. Background

In March 2020, Congress passed the Broadband DATA Act (BDA). BDA required the FCC to "create a common dataset of all locations in the United States where fixed broadband internet access service can be installed, as determined by the Commission." The purpose of the dataset is to "serve as the foundation upon which all data relating to the availability of fixed broadband internet access service collected... shall be reported and overlaid".[3]

The Broadband Serviceable Location Fabric meets these requirements.

Before the passage of BDA, CostQuest Associates worked with a group of Internet Service Providers (ISPs) to implement a prototype project to investigate the viability of producing a broadband fabric.[4] The pilot project was designed to answer the following four questions:

1. Could a data model for broadband serviceable locations be developed and populated with commercial or open-source datasets?
2. Was it possible to identify locations requiring broadband service?
3. Was it possible to match geographic and/or address-based Internet Service Provider (ISP) broadband service provisioning information to the serviceable location?
4. Was it possible to enumerate broadband coverage at a serviceable location level?

[1] Broadband Deployment Accuracy and Technological Availability Act, Pub. L. No. 116-130, 134 Stat. 228 (2020) (codified at 47 U.S.C. §§ 641-646) (Broadband DATA Act).

[2] The Fabric is developed in two distinct steps. The first is referred to as the Fundamental Fabric. This refers to the development of the geospatial attributes (latitude and longitude). This second step involves development of textual attributes like address, unit counts, bsl_flag. This distinction is discussed in section 2.2.2.

[3] *ibid*

[4] When Fabric or Broadband Serviceable Location are capitalized in this document it refers to the Broadband Serviceable Location Fabric as designed by the FCC to support the Broadband Data Collection. When not capitalized, it refers to earlier implementations, not compliant with FCC direction to support the Broadband Data Collection.

Project sponsors selected Missouri and Virginia for investigation.[5] The project team first investigated source datasets. Next, the team developed a preliminary data model. Finally, software developers designed, tested, and implemented geoprocessing techniques.

The team produced an initial broadband fabric for both states.

CostQuest received coverage information from ISPs and associated their coverage data with a serviceable location. With the association of ISP coverage information to a serviceable location, it was possible to identify areas lacking coverage at a more granular level than the Census block standard used in the FCC's Form 477 collection. In summary, the pilot project demonstrated the technical feasibility of broadband fabric development and the benefits derived from the granular information.

1.2. Identifying Serviceable Locations

During the pilot project, we learned that identifying a broadband serviceable location was at times idiosyncratic and highly dependent on the quality of the data available, the variables considered, the models implemented, and manual intervention. When reviewing potential data sources, e.g., building footprints, parcels, and addresses, it became apparent that there was no single, standard data source or method to identify broadband serviceable locations. Additionally, early project development faced the hurdle that the identification of a broadband serviceable location was partially subjective and not wholly deterministic.

The pilot project showed that identifying a broadband serviceable location required complex logic-based determinations based on training datasets developed by our internal broadband experts. In reviewing potential data sources, including building footprints, parcels, and addresses, it was apparent that:

1) no single data type listed above could, in isolation, definitively identify a serviceable location;
2) there were no known methods in place to use the discrete data sources to identify broadband serviceable locations; and,
3) training[6] datasets were sparse, and the identification of broadband serviceable locations was subjective and not a defined process.

The development team addressed these issues during the analysis and design phase of the pilot. The third issue, the lack of a standard method for broadband serviceable location identification, prompted software developers to implement an architecture similar to an Expert System.[7] In effect, CostQuest used a team of experts to review the combination of source datasets and identify where a broadband serviceable location existed. During this early work, they established a teaching flowchart, shown in Figure 1. The flowchart was used across the project team so that people reviewing potential broadband serviceable locations would reach consistent decisions.

[5] A presentation describing the methods and findings of the pilot project is available for review https://energycommerce.house.gov/sites/democrats.energycommerce.house.gov/files/documents/Stegeman_Appendix%20A.pdf

[6] A training dataset refers to a dataset of known accuracy. The training dataset is used to inform a computer algorithm with the goal that the algorithm can use the training and apply the findings to an external dataset to reach valid conclusions. Training datasets are used in a number of fields including Artificial Intelligence.

[7] An Expert System is a computer application that mimics human decision making. See for example, https://hbr.org/1988/03/putting-expert-systems-to-work

The project team designed a Quality Assurance (Q.A.) process that compared and quantified the reviewers' results. The QA process supported measurements of confidence in a particular serviceable location using multiple techniques. It also allowed for measures of inter-reviewer reliability.

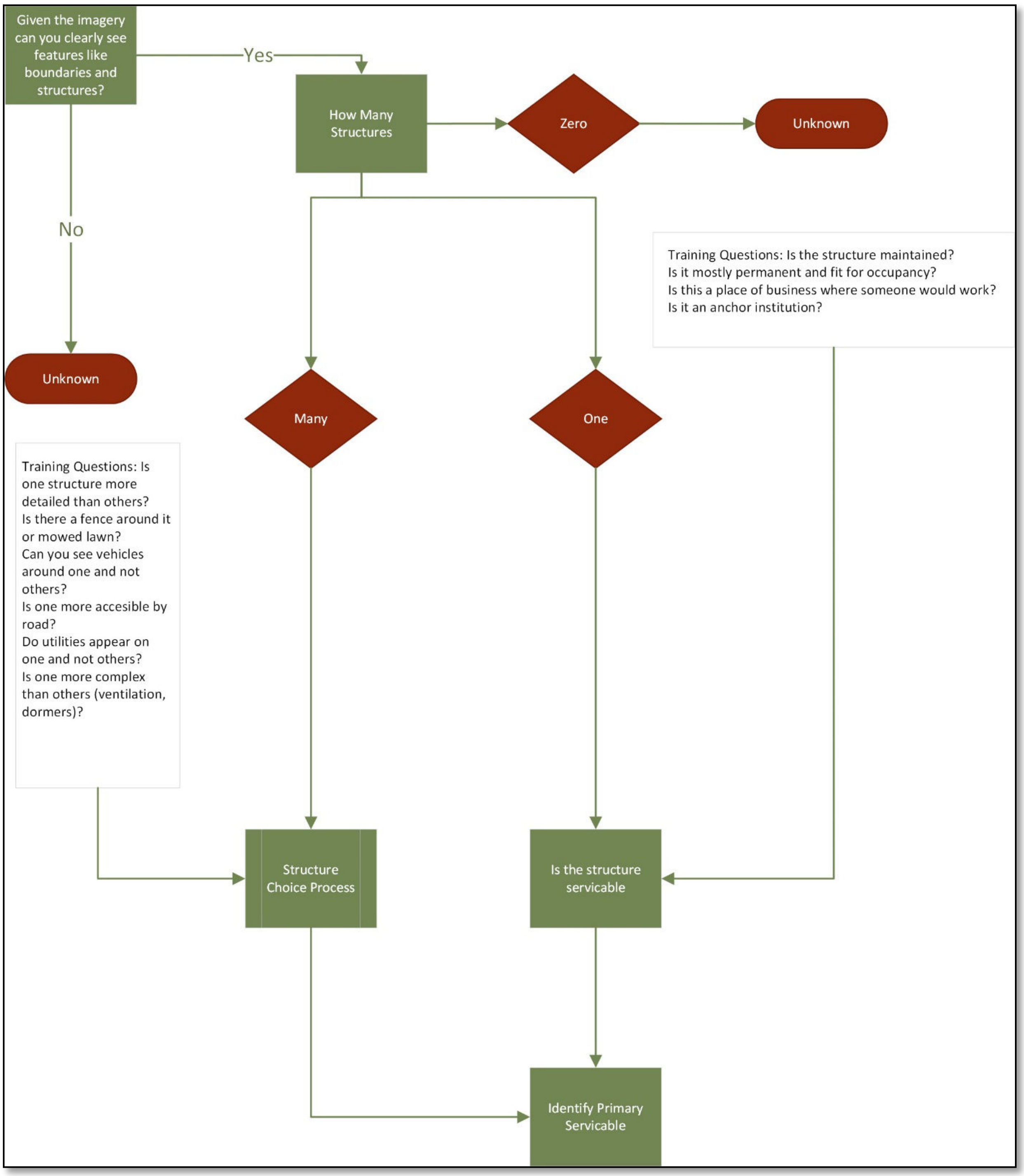

Figure 1

Over time, the review team developed a knowledge base of combinations of input data sources and review findings. Analysts studied the relationships among data sources and expert review. Statisticians then proposed models to reflect similar review conclusions. Analysts compared the performance of each model against the known conditions identified in the knowledge base.

1.2.1. The Role of Input Data

CostQuest receives data sources in different formats from multiple vendors. Each dataset is designed for a specific purpose and delivered on a schedule independent of other datasets. This means that the Broadband Serviceable Location Fabric (Fabric) production process must deal with heterogeneous and sometimes inconsistent data sources. Some examples of these inconsistencies are described below:

- Although standards exist for sources like textual addresses and parcel geometry, there is no guarantee that every jurisdiction will follow these standards.
- Information received in one source may conflict with another. For example, imagery may not show a structure, but assessment information may reflect significant improvement value.
- Vendors release datasets on their calendar, not synced to a Fabric release. A Fabric ground date will reflect the release dates from many vendors.
- Property assessment and parcel boundary information are not always available. There are some counties[8] for which parcel polygons are not available. In addition, there is a significant amount of non-parceled land.
- In the circumstances like condominiums where ownership of the land is distinct from ownership of a structure, counties implement different data models to convey this distinction. CostQuest's models attempt to standardize this circumstance, but the presentation of overlapping and duplicate geographic datasets makes geoprocessing complex.
- County boundaries are not always consistent. Counties may maintain information for locations outside their county, or their designation of the county boundary differs from their adjacent neighbor or the Federal government
- Counties may not update their assessment information or may update as needed. This can make information correct in some places and incorrect in other portions of the same county. In addition, assessment officials may update some attributes—like a tax billing address more frequently than a situs address.
- Source vendors develop data sources with different levels of spatial accuracy. This can be seen within the structure use datasets. We have found most of the input data is developed to a high standard. In other places, there is a noticeable location error.
- Spatial conversions can impact accuracy. This type of issue can occur when a paper data source is registered incorrectly prior to conversion to digital format.[9] Fabric data

[8] The Fabric is produced on a county or county equivalent basis. There are cases where the assessing jurisdiction is not a county or county equivalent, but the term county is used to represent the set of assessment data received for the county or county equivalent.

[9] Registration is the process of associating an image with geographic reference like longitude and latitude, see: https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/register-raster.htm

processing is a highly geospatial exercise. Results are impacted negatively by imprecise and/or different projections into an X.Y. (planar) coordinate system.

A significant portion of the BSLF development process involves standardization and clean-up of the input data sources. Even after these processes, there are cases where source inconsistencies are irreconcilable and conflicting information is presented. In these cases, we work on improving the underlying data (e.g., new aerial imagery acquisition) and use correction data from various sources to address these issues.

1.3. Purpose-Driven Development

The BSLF conforms to the criteria set forth by the FCC. Specific characteristics of Fabric development, data architecture, and deployment meet and support identified criteria.[10] These key design attributes were set by the FCC as follows:

- A BSLF record is a non-residential or residential location in the United States at which fixed broadband internet access service (i.e., mass-market broadband) is, or has the potential to be, installed by an ISP; it is a single point defined by a set of geographic coordinates that fall within the footprint of a building.[11]
- Residential and non-residential locations are defined based on the assessor's land use indication, the United States Postal Service's delivery identification, and additional datasets that identify Community Anchor Institutions and group quarters such as military installations, jails, and college campuses.[12]
- A Multi-Tenant Environment (MTE) is a single location record in the Fabric with more than 1 residential or business unit. The number of units within each MTE should be developed as an attribute to that Fabric data record.
- Locations such mobile home parks, marinas, and other transitory locations reflect broadband serviceability to the extent determinable.
- Mixed business/residential locations (e.g., multi-tenant units with commercial space on the ground floor and apartments above) reflect broadband serviceability to the extent determinable.

Due to these design criteria, the FCC BSLF is a unique product and differs in structure and presentation from other location datasets. For example, an E911[13] dataset may have similar characteristics to the FCC BSLF; however, the FCC BSLF represents the location/connection point for broadband service, while the E911 dataset represents where emergency responders should respond to. Many of these location records in each source overlap, but the datasets are not identical. For example, the FCC BSLF will represent an MTE as a single point, while an

[10] FCC Public Notice, DA 22-413

[11] There are cases where we have incomplete or contradictory data with available imagery lacking a building, but other data sources indicate significant improvement on the land. In such situations we will create a Fabric point approximating the location of the building. In other cases, the precise building footprint may be somewhat indeterminant due to shadowing, adjacent structures, etc.

[12] Specific requirements for Community Anchor Institutions (CAIs) and Group Quarters (GCs) fulfill Commission requirements as described in later sections. See Table 2 for Data Sources, Implications of Structure Use—Entity Boundaries and Development of the Building Type Code, sections.

[13] Enhanced 911 is abbreviated as E911, see https://www.fcc.gov/general/9-1-1-and-e9-1-1-services.

E911 dataset may represent each living/housing unit distinctly or identify each distinct building entrance. Another example of a similar dataset is address data. A point-based (spatial) address database may list the approximate locations of delivery points. But it may not contain information about the type of location being delivered to (business, residential or mixed customer) or even if a structure exists to be delivered to. For example, delivery points may represent empty parcels, loading docks, mail delivery drop boxes, or other non-serviceable locations. The address dataset may also exclude locations that are not commercially deliverable.

Each source dataset used in the creation of the Fabric is important and useful; however, they are different and, in some cases, inconsistent with one another, given their different purposes. The BSLF must address these differences and inconsistencies and synthesize the source datasets in the best way possible to meet the requirements defined by the Broadband DATA Act and the FCC.

1.4. Quality and Reliability Measures

The BSLF represents the best available source for identifying Broadband Serviceable Locations. One of the primary strengths of the product is the uniformity with which it treats and interprets conflicting source datasets.

With that said, BSLF accuracy is impacted in areas where there is a lack of uniform source datasets or the licensing requirements for available datasets prevents their use in fulfilling the BDA. Areas that lack standardized addressing,[14] standardized systems for land management and conveyance,[15] up-to-date imagery and/or reasonable commercial data usage policies may therefore vary in terms of BSLF accuracy. Given this, there are two important considerations. First, there are some regions of the United States and its territories where the best mechanism for a provider to match BSLs is a geographic search instead of text searches (i.e., address geocoding). Second, the focus of BSLF product development has been to establish a baseline for serviceable location identification and attribution and, from that baseline, allow successive iterations complemented with stakeholder challenges and improvements. The importance of the BSLF community is discussed in the next section.

1.5. Supporting a User Ecosystem

The FCC BSLF does not exist in isolation. It is an important component of multiple FCC efforts. As such, beyond data processing, dataset availability, and implementation of business rules, it was equally important to create methods and procedures to support the distribution, correction, and enhancement of the BSLF. In this document, we use the term ecosystem to refer to the Fabric processes and functions related to and supporting the BDC. The ecosystem is as important as the presence of the Fabric. Key aspects of the ecosystem include the following:

1. Fabric data files are consistently produced and accessible for interrogation and review. For the FCC BSLF, data files are visible and available in system-neutral and human-readable Comma Separated Variables (CSV) format.

[14] This article discusses the non-uniform address situations in Puerto Rico, https://www.wired.com/story/geocode-address-puerto-rico-hurricane-maria/. HUD describes similar complexities, see https://www.huduser.gov/portal/pdredge/pdr-edge-frm-asst-sec-070918.html The Native American Rights Foundation also describes address and postal delivery issues on Tribal lands, https://www.narf.org/vote-by-mail (visited June 6, 2022)

[15] Lands designated as tribal can have a mix of ownership statuses. This complex ownership may make assignment of parcels to owners and addresses very complicated. See https://sgp.fas.org/crs/misc/R46647.pdf

2. The Fabric data files must be presented in a format easily ingested by commercial and open-source platforms and Geographic Information Systems (GIS) applications.
3. The Fabric data files should be made available in such a way as to support challenge, correction enhancement, and modification.
4. The Fabric is designed such that the BSL location record[16] remains persistent version to version when newer evidence indicates the position of the serviceable location or the presence of a serviceable location on a single location parcel is not significantly changed as compared to the prior version.
5. Fabric data files are accessible to a large community of users. Therefore, licensing mechanisms must support a variety of user types (ISPs, Government, Third Parties), and end-user licensing must ultimately support the BDC.

2. FCC BSLF Vocabulary and Production Architecture

This section discusses fundamental BSLF concepts. To clarify concepts and vocabulary, we will introduce a series of images. To develop the images, CostQuest transformed the Fabric data file using a Geographic Information System (GIS). Each image provides readers with visual context for terms used within the Fabric data model. A later section provides an overview of BSLF output development.

2.1. BSLF Vocabulary

In this section, we clarify key terminology used in this documentation.

Figure 2 shows the transformed BSLF output displayed over visible imagery.

[16] Each location record in the Fabric data file is identified with a location_id. The location_id is the primary key for the Fabric data file. The location_id will remain persistent to indicate there is not a significant change in the position of the BSL.

Figure 2

Broadband Serviceable Location Fabric (BSLF): a geospatial dataset that identifies locations that receive or could receive fixed-broadband Internet access service from an internet service provider. When populated, the database consists of location records that either receive or are candidates for mass market or non-mass market broadband service in the United States.

In the Fabric production process, all potential structures are identified. Features that are too small or don't conform to modeling standards[17] are excluded from consideration. These are considered supplementary records and excluded from the Fabric data file.

The BSLF data model classifies location records as Inactive (red points) or Active (green points). An Active location record qualifies as a potential BSL. An Inactive location record represents a unit of area (a polygon) that is identified but not classified as Active. Inactive records are typically area placeholders. They capture areas identified as vacant land, parks, forests, and cropland. Inactive locations are not available in the Fabric data file. Each Active location record is classified as BSL Flag TRUE or FALSE.

Broadband-Serviceable Location (BSL): an Active location record that does or can receive mass-market broadband internet access. Each location record is described by various attributes such as a unique location_id, address, unit count, land-use, or BSL Flag (bsl_flag). Fabric data files retain all BSLs. Records marked with the bsl_flag attribute of TRUE are BSLs. Non-BSL

[17] These rules work to exclude features like sheds, garages, guard sheds, parking lot attendant buildings, out-buildings, etc. These are not considered *locations*, either mass market or non-mass-market.

records, including location records that we expect would not subscribe to mass-market broadband service, are marked with the bsl_flag of FALSE.

Figure 3 zooms into the prior image. In reviewing this map, several other descriptive terms become important.

Figure 3

Parcel: a parcel is a polygon representing an area of common land ownership.[18] The red lines in Figure 3 represent parcel polygons. Parcel data is not included in the Fabric, but instead is used as a constraint to inform the processing logic by geographically categorizing other unrelated data. A parcel itself can be categorized as either a Single Location Parcel (SLP), which means there is only one potential BSL on a parcel, or a Multi-Location Parcel (MLP). Most parcels are SLP (e.g., single-family home on parcel). Multi-Location Parcels tend to be cases like shopping centers, apartment complexes, or office complexes. There are also cases of Multi-Location Parcels where a parent/child parcel relationship exists. A parent/child parcel relationship means a set of parcels are entirely contained within another parcel. For example, some jurisdictions[19] use parent/child parcel relationships to represent the case where a condominium structure is owned by a different entity than the housing units within. This relationship also exists within

[18] In development of the BSLF, we focus on Parcel boundaries or polygons. The International Association of Assessing Officers (IAAO) defines a Parcel as: a contiguous area of land described in a single legal description or as one of a number of lots on a plat; separately owned, either publicly or privately; and capable of being separately conveyed, see: https://www.iaao.org/media/Pubs/IAAO_Glossary.pdf.
[19] Jurisdictions use multiple standards to represent land and structure ownership. There appears to be no single standard for parcel polygon development or relationships among parcels. Because there exists no single standard,

developing models to interpret parcel relationships and geometries to inform the logic is extremely complicated.

some residential communities where a homeowner's association owns common areas, and those common areas overlap other parcels.

Footprint: a footprint is a polygon extracted in an automated process from remote-sensed data that represents the shape or bounds of a structure. In the image above, the orange polygon represents the extracted footprint. The footprint approximates the shape of the remotely sensed object. Footprints are extracted from orthorectified visible imagery and LIDAR point clouds.[20]

Address: an address is a human-readable attribute to identify a location. The text string that composes the address is rarely directly transferable to a geographic location, such as longitude and latitude. An address can vary in form for different purposes. For example, an emergency routing address versus a United States Postal Service (USPS) address versus an alternative carrier's address may have dissimilar textual variations describing the same geographic location. In some cases, an address may be geographically precise when it is developed as a reference to nearby roads (e.g., the intersection of CR 232 & Hwy 173) but may be challenging to standardize and interpret at scale for the BSLF data model. An address may also exist for purposes other than structure identification. Utility equipment may be addressed by the equipment owner for their equipment inventory purposes. Land that is vacant but parceled may have an address. Sometimes this address is unique, but in other cases, it is established based on the address of an adjacent parcel. Structures on private or unimproved roads and structures not receiving USPS or commercial deliveries may not be addressed or may have non-location specific addresses like P.O. Box or Highway Carrier Route references. For the reasons mentioned above, a textual address is not always specific to a geographic location, and not all serviceable locations have addresses.

Within the BSLF data model, an address is characterized as either primary or secondary. A primary address is the standardized[21] form of an address that the data model interprets as the most likely address for a record. A secondary address is an alternative address that refers to the same geographic location. This could be a building that faces two streets and has been assigned more than one address by an assessor. Because a road may have multiple names, there are cases where a secondary address for a location has a different street name than the primary address. The secondary addresses assigned in our process are not an exhaustive list of all secondary address permutations.

Before closing the terminology discussion, a few other aspects of the BSLF data model require additional clarification.

First, a record in the Fabric output file can represent zero to many identifiable structures. Although most are a single structure, there are cases where a BSL may represent multiple

[20] Orthorectification converts images into sources suitable for maps. LIDAR is a method of remote sensing relying on lasers rather than visible light for remote sensing.
[21] Standardized relates to how the address is parsed apart into components. Because not all addresses are deliverable (e.g., private roads, non-delivery communities, infrastructure) by commercial carriers some addresses present may not be USPS deliverable. Not all addresses presented may be in USPS recommended form.

structures (see, e.g., section 0, below). There are other cases where a BSL may be present, but imagery conflicts with other data sources and evidence of a structure is inconclusive.

Second, a location record may or may not be in a parceled area. There are cases where land is not parceled, as well as cases where a single structure can occupy multiple parcels.

Third, although some address data sources may represent the same textual address at multiple geographic locations, the BSLF data model imposes a restriction that a textual address may appear in one and only one active location record. In the case of an entity boundary--a college/university, a military reservation, or a jail-one to three location records are presented in the Fabric data file. Each location record for the entity boundary records has a non-duplicate address.

Fourth, a location_id uniquely identifies a location record. The location_id is a persistent identifier.

2.2. BSLF Output--Overview

Developing BSLF output is a multi-stage process. Although each step is described in detail in later sections, we provide a brief introduction guided by Figure 4, below.

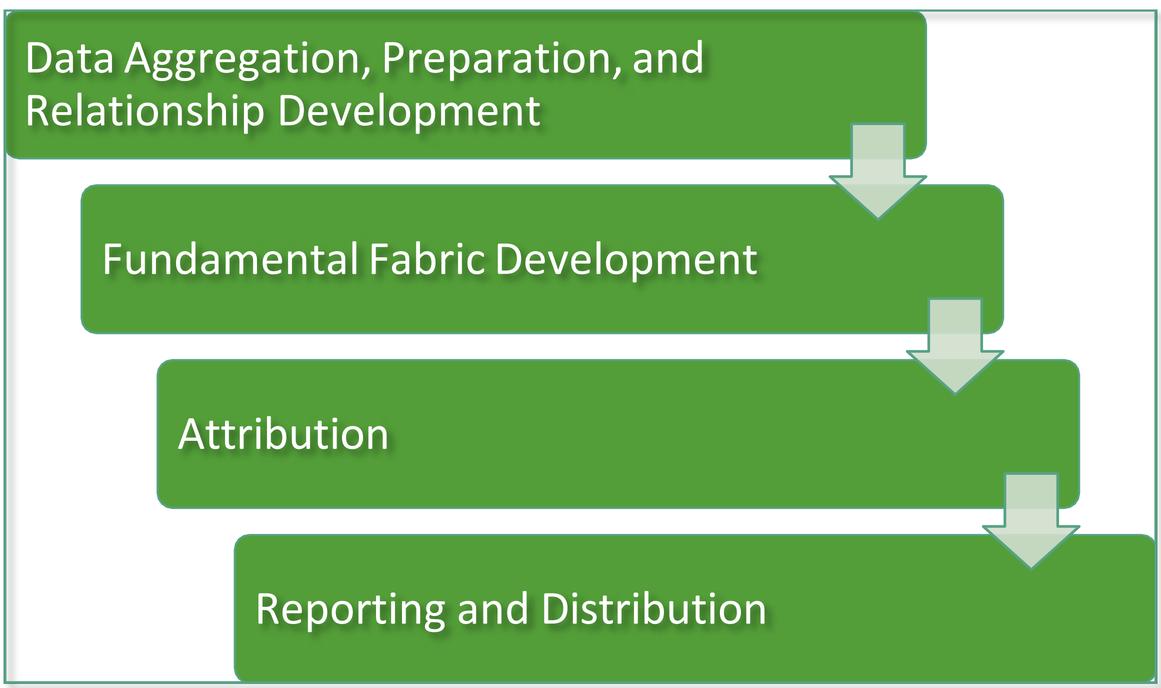

Figure 4

2.2.1. Data Aggregation, Preparation, and Relationship Development

The first preparation step involves loading source data into project databases, establishing initial quality measures, and writing internal metadata. The metadata identifies each dataset, its source, and when it was loaded. Each dataset is standardized in format as well as geographic structure.

The process examines each discrete source and develops relationships among the various sources. Finally, the data sources are ranked, and discrete measures of confidence in the quality of each data source are established.

2.2.2. Fundamental Fabric Development

Fundamental Fabric development involves a sequence of steps that import data, establish relationships among data sources, execute models to assess the likelihood of a BSL and then characterize BSLs. The endpoint of Fundamental Fabric development is identifying serviceable locations and a set of address candidates. The phrase Fundamental Fabric refers to the intermediate stage in BSLF production before the final address, and other attributes are attached to a location record.

2.2.3. Attribute Development

There are several processes involved in attribute development. The first process identifies the nature of the building type (residential, non-residential, mixed, etc.). The second process develops the count of units within each location record. The third assigns an address. The final process examines the information and updates the bsl_flag to be TRUE or FALSE.

2.2.4. Reporting and Output Distribution

After completing the steps above, CostQuest prepares BSLF data for output. This involves the development of user-specific data packages, file validity checks, posting to a secure website, and distribution to licensed users. Finally, a separate data feed is presented to the FCC for deployment into the BDC system.

3. BSLF Production-Detail

As described earlier, BSLF production consists of distinct processes: Data Aggregation Preparation and Relationship Development, Fundamental Fabric Development, Attribution, and Reporting and Distribution. Each process is dependent on the data presented to it, as well as underlying sub-processes.

The following sections discuss each step. First, we describe the data sources and then explain how the data sources support the production process.

3.1. Data Sources

An objective of the Fabric data model is to ingest multiple datasets. Multiple source datasets are necessary because no single existing data source meets the unique requirements of the BSLF. In addition, multiple data sources can corroborate key details or supplement information when a primary data source has a gap. In sum, using multiple data sources produces a final product that is more consistent with BSL definitions than any source datasets independently.

Two categories of input data sources are used in Fabric production. The first set represents sources that develop the Fundamental Fabric. This set is described in Table 1. The second set provides input to the structure use process. The structure use process helps set the bsl_flag value.

3.1.1. Fundamental Fabric Data Sources

We describe source datasets in Table 1 below. The source and vintage (load date) are included for reference.

Data Element	Description
Address Datasets	Address data provides textual fields[22] to associate a footprint with a textual version of the geographic location
Address Validation	Records provided by address data vendor are provided in a standardized form. Coding Accuracy Support System (CASS[23]) certified software is released on a periodic basis and is used to validate addresses for reduced mailing rates.
Assessor Information	Tabular information is used for internal processing only. The assessor data represents attributes of ownership and/or value. This may include address, value of buildings, number of buildings, number of units, land use, etc.
Carrier Ingested Data	Location information submitted by serving carriers
U.S. Census Data	Geospatial data developed by U.S. Census. TIGER Line files are vector files developed by the U.S. Census
Land Cover	Polygon data containing generalized land use information (impervious, agricultural, industrial, etc.)
Parcels	A geospatial polygon. The parcel boundaries provide an area of land ownership. They are used as a constraint to inform the model logic as to how to associate other unrelated data such as assessor information on valuation, land use, and structure characteristics. Parcel Boundary data is not itself included in the Fabric.
Roads	A national dataset of road centerlines in GIS format. Contains corresponding information regarding street names, alternate names, address ranges, road types.
Structure Footprints	Feature extraction algorithms use visible imagery or LIDAR point clouds to identify building footprints.

[22] Many sources remove diacritical marks (accents, tildes, etc.) from source address data.
[23] Coding Accuracy Support System, a USPS certification of software that analyzes addresses and rates its quality. CASS software can also parse and standardize addresses.

Data Element	Description
Unit Counts	Address-based information for occupied housing units and businesses. Can characterize types of location (residence, business, enterprise) and delivery location.

Table 1

3.1.2. Structure Use Data Sources

We describe each structure use data source in Table 2. The source and vintage (load date) are included as well.

Structure use data sources flow into the Fundamental Fabric in two ways. First, structure use polygons representing specific types of structures, such as mobile homes, inform CostQuest models on the treatment of footprints in a particular area. Second, the presence of specific structures (geographic coordinates) informs the decision to indicate if a location record is a BSL (bsl_flag equal to TRUE).

The BSLF development process receives structure use data in both point and areal forms, described in Table 2 below.

Structure Use Category	Data Type	Illustrative Source	Identifies	Vintage
Special Access/ Enterprises	Business Type and Employment Estimates	Business information data vendor	High Bandwidth Business Locations	Extract April 2022
Government	Courthouses	https://hifld-geoplatform.opendata.arcgis.com/datasets/courthouses	CAI (bsl_flag=FALSE if in CDP[24] with >= 500,000 Population)	Extract November 2021
Government	EMS	https://hifld-geoplatform.opendata.arcgis.com/datasets/emergency-medical-service-ems-stations	CAI (bsl_flag= FALSE if in CDP with >= 500,000 Population)	Extract November 2021
Government	Fire	https://hifld-geoplatform.opendata.arcgis.com/datasets/fire-stations	CAI (bsl_flag= FALSE if in CDP with >= 500,000 Population)	Extract November 2021
Government	Governor Mansions	https://hifld-geoplatform.opendata.arcgis.com/datasets/governors-mansions/explore?location=38.405747%2C-110.396312%2C3.84	CAI (bsl_flag= FALSE if in CDP with >= 500,000 Population)	Extract March 2022
Group Quarters	Group Quarter Population Counts	U.S. Census, Housing Units PL-194 extract 2020	Group Quarter	Extract U.S. Census 2020
Healthcare	Hospitals	https://hifld-geoplatform.opendata.arcgis.com/datasets/hospitals	CAI	Extract November 2021
Law Enforcement	Local Law Enforcement	https://hifld-geoplatform.opendata.arcgis.com/datasets/local-law-enforcement-locations	CAI (bsl_flag= FALSE if in CDP with >= 500,000 Population)	Extract November 2021

[24] CDP refers to a Census-Designated Place.

Structure Use Category	Data Type	Illustrative Source	Identifies	Vintage
Government	Major State Government Buildings	https://hifld-geoplatform.opendata.arcgis.com/datasets/major-state-government-buildings	CAI (bsl_flag= FALSE if in CDP with >= 500,000 Population)	Extract March 2022
Military	Military Bases (National Boundaries / Military Reserve)	https://apps.nationalmap.gov/downloader/	CAI, Group Quarters, Units	Extract November 2021
Modeling	Mobile Home Parks	https://hifld-geoplatform.opendata.arcgis.com/datasets/mobile-home-parks/explore?location=41.170276%2C-114.616034%2C4.07	Mobile Homes	Extract November 2021
Law Enforcement	Prisons	U.S. Census Geographic Features	Group Quarter	U.S. Census 2020
K-12 Education	Private Schools	https://hifld-geoplatform.opendata.arcgis.com/datasets/private-schools	CAI	Extract November 2021
Healthcare	Public Health Dept	https://hifld-geoplatform.opendata.arcgis.com/datasets/public-health-departments/explore?location=36.403361%2C-106.164623%2C3.83	CAI	Extract November 2021
Library	Public Library Branches	https://www.imls.gov/sites/default/files/2021-05/pls_fy2019_csv.zip	CAI	Extract November 2021
K-12 Education	Public Schools	https://hifld-geoplatform.opendata.arcgis.com/datasets/public-schools	CAI	Extract November 2021
Healthcare	Red Cross Facilities	https://hifld-geoplatform.opendata.arcgis.com/datasets/american-red-cross-chapter-facilities	CAI	Extract November 2021
University and college	Supplemental colleges	https://hifld-geoplatform.opendata.arcgis.com/datasets/supplemental-colleges	CAI	Extract November 2021
University and college	University and colleges	U.S. Census Geographic Features	CAI, Group Quarters, Units	Extract U.S. Census 2020
Healthcare	Veteran's Administration Facilities	https://hifld-geoplatform.opendata.arcgis.com/datasets/f11d7d153bfb408f85bd029b2dac9298_0/explore?location=35.446976%2C-107.008404%2C3.62	CAI	Extract November 2021

Table 2

3.2. Production

We implement the BSLF production process across the steps shown in Figure 5. The summary steps on the top row are composed of significant sub-processes.

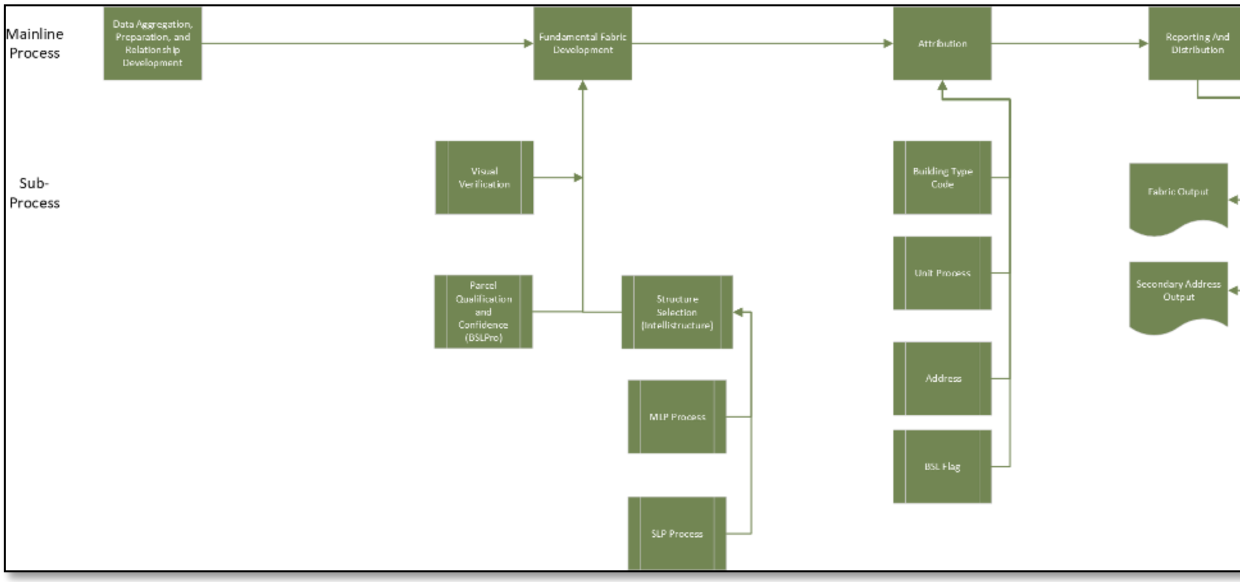

Figure 5

The remainder of this section describes BSLF production.

3.2.1. Data Aggregation, Preparation, and Relationship Development

The Data Aggregation, Preparation, and Relationship Development processes perform extensive work to standardize source data. Some of this work involves geographic topology clean-ups. Although a structure footprint or parcel polygon may be available, there is no guarantee that the polygon was developed in a valid manner. Invalid polygons can create unexpected processing results. [25] Hence, extensive effort is expended to make sure all geometries presented are valid. In addition to polygon validity, the process needs to unstack geometries, fix overlaps and gaps, remove duplicates, and ensure unique identifiers for distinct ownership information and structures. While parcel data arrive from our data vendors with a property identification, they are not linked to the tax assessor records. We have created an enhanced linkage process to join these two datasets using a persistent identification field that is unique to each property record. However, there are some parcel records that are not linked to a corresponding tax assessor record. This impacts the downstream models that analyze the footprint records because fewer informative variables are available. The enhanced linkage process reviews records and identifies edge cases in which additional records could be linked using an intelligent textual clean-up process that looks across multiple potential columns for matches. The processes described above create relationships linking one layer to another. Relationships among records within each layer are also important, and the process identifies, for example, parcels that are contained entirely within other parcels as a layer of sub-parcel data. Each source data type requires its unique approach to cleaning and linking to other layers.

Parcel Qualification. In many cases, parcel geometries require clean-up to be usable to inform the processing logic. Fixing geometries so that they do not overlap with one another, as well as removing parcels that represent catch-all areas within the road network of a city or county, are

[25] Our process utilizes PostGIS and makes use of OGC compliant functions to validate geometry. See https://postgis.net/workshops/postgis-intro/validity.html (visited 09/25/2022)

necessary. Jurisdictions do not apply a single standard for parcel data, which leads to a wide variety of data presentation choices. Certain jurisdictions may provide parcels that are duplicates and stacked on top of one another in cases where there is one parcel that has multiple units, whereas others may provide one parcel and indicate the number of units. It is not uncommon in urban areas to see individual units carved out of a larger parcel containing it. Certain parcels represent individual units such as townhomes or businesses in a strip mall. Our process identifies and normalizes these differences into a consistent representation to advance the modeling process.

Structure Qualification. Structure footprints require less clean-up and validation than parcel geometries. Yet there is some variation in how vendors provide their data that must be accommodated when aiming for consistent modeling results. Some datasets will provide buildings split by parcel boundaries to represent units more accurately in cases of townhomes and commercial strip malls, while others may not. In rare cases, structure geometries may need small clean-up steps to be considered valid for geoprocessing purposes. This step also performs a preliminary identification of structures that are highly unlikely to be serviceable based on attributes such as their size and shape. This removes structures, such as grain bins and silos, as well as extremely small "noise" structures, such as sheds.

3.2.2. Fundamental Fabric Production

Fundamental Fabric production processing begins after data have been linked, cleaned, and standardized. Given the input data source, Fundamental Fabric production answers three questions:

- Does this parcel contain a Broadband Serviceable Location?
- When looking at an area of land, which structure should be selected as the Broadband Serviceable Location?
- Given the uncertainties in the modeled selection process, is there any information in an override data source that should change this decision?

CostQuest's BSLPro model answers the first question. The next question is answered with CostQuest's Intellistructure model. The final question is answered with CostQuest's visual verification database.

CostQuest developed independent models to identify Broadband Serviceable locations. BSLPro is a modular logit-based regression model and continues to evolve as more data informs it. During development, the statistical model was benchmarked against various machine-learning-based approaches and determined to be similar in accuracy. Intellistructure is a series of logistical regression (logit) models. CostQuest calibrates both models using a training dataset sourced from expert reviewers internal to CostQuest. The dataset is updated for major releases to incorporate additional records. For a record to be considered adequate for use in the training data, that record requires confirmation from multiple expert reviewers.

3.2.2.1. BSLPro

BSLPro assigns a confidence score to every parcel representing the likelihood that there is (or is not) at least one Broadband Serviceable Location contained in an area. BSLPro outputs are superimposed on imagery in Figure 6

Due to the disparate availability of data across the country, there are two types of modeling. The first model works where a link between the parcel and tax assessor records can be created; the second where parcels are lacking or have not been linked to tax assessor data. Dual models allow for the maximum degree of confidence in results. Both models consider more than a dozen characteristics when developing the score. These include the profile and characteristics of every structure contained on the parcel, land use, land cover, census indicators, structure parameters from the tax assessor, and more. BSLPro is highly configurable and evolves to become more accurate as more data is obtained and more records are verified as Broadband Serviceable or not. Currently, it is calibrated to more than 25,000 randomly selected records. Random selections were made from a stratified dataset to avoid any bias in sampling.

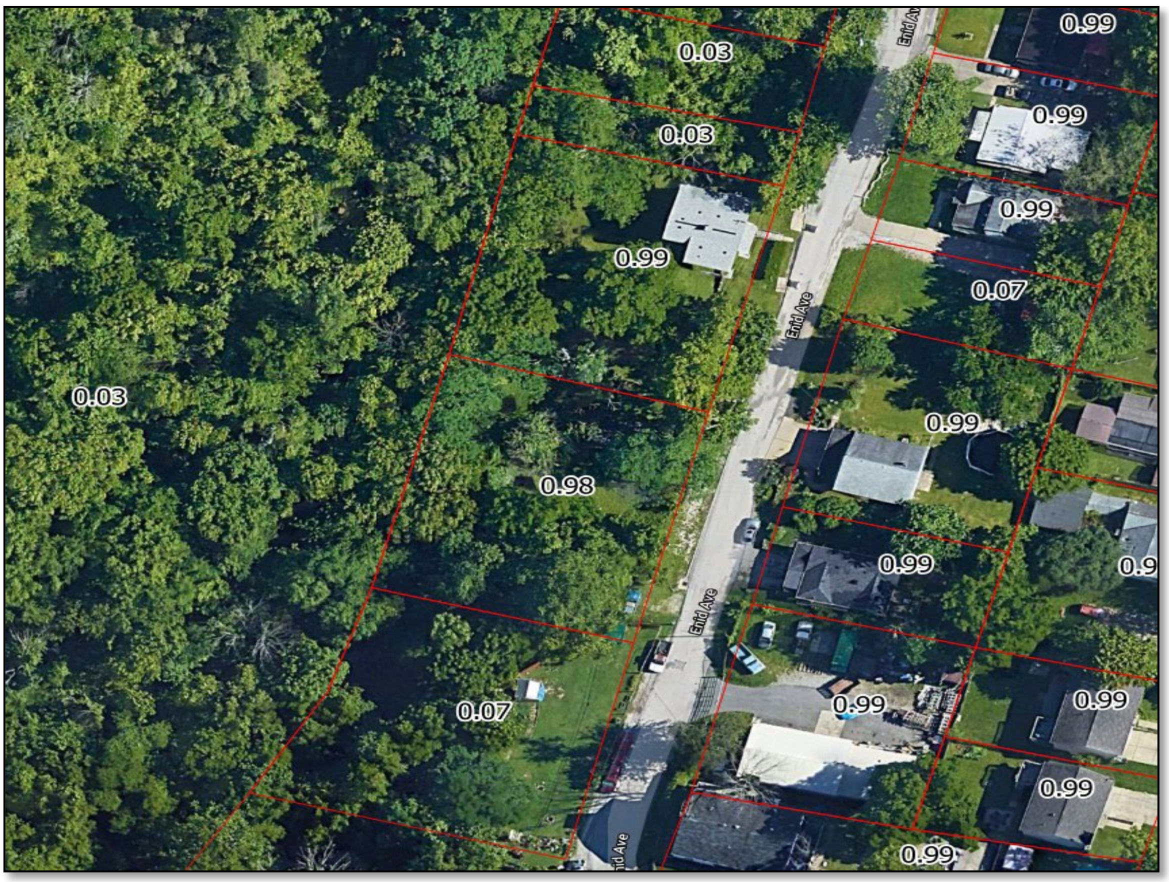

Figure 6

3.2.2.2. Intellistructure

After BSLPro identifies whether an area is likely to contain at least one serviceable location, Intellistructure interrogates the collection of structures on the parcel, as represented in Figure *8*. Intellistructure is fit to a set of data found on parcels with multiple structures. These records are sourced from a sampling of multi-structure parcels across a variety of land use and land cover conditions. Both expert and crowd-based reviews are combined to create the calibration dataset.

Intellistructure utilizes more than ten unique characteristics to determine those likely to require broadband service. In general, these characteristics include profiles of the structures to be considered, such as their size, shape, and complexity.

Figure 7

Intellistructure creates its score and determinations based on on an extensive set of attributes, including:

- structure footprint characteristics
- distance to largest and smallest structures
- structure height and roof slope
- next smaller and larger structure characteristics
- land area characteristics
- number of structures
- whether the parcel is identified as a likely multi-location parcel
- tax assessor attributes

Implications of Multi-Location and Single-Location Parcels

A single serviceable location for a parcel is determined when the source data indicates single-family, residential or when there is an identified single housing structure. Fundamental Fabric production then identifies multiple serviceable locations among remaining parcels where there is more than one residential building with single or multiple housing units. In such cases, each building on the parcel may be considered a serviceable location.

Multiple serviceable locations are also identified on a non-residential parcel if the parcel contains multiple commercial, office, or industrial buildings (but not parcels that are identified as agricultural). In such cases, buildings identified by the Intellistructure model would be considered serviceable. This would include business parks with multiple structures, each with its own occupants or tenants. If data indicate the locations are occupied by a single tenant, the unit count for the structure will be set to 1.

Implications of Structure Use—Entity Boundaries

Where there is structure use data (see section 3.1.2) indicating the presence of an entity such as a military base, a prison, or a college or university, the entity's boundary overrides the parcel boundaries, and a single representative serviceable location is selected within the entity boundary. In the case of an entity boundary, one to three location records are created. We create a location record with a building type code (building_type_code) of "C" (Community Anchor Institution). If Census data indicate the presence of housing within the entity (excluding dorms and barracks as Group Quarters), we create a second location record (building type code, R). Finally, if the entity area information indicates the presence of Group Quarters, we create a third location record (building type code G).

Non-Parceled Areas

All land area is not parceled. Figure 9 shows large non-parceled areas to the East and West of the parceled areas. In the image, parcels are symbolized with red lines.

Figure 8

In areas without parcels, an alternative model is used. This no-parcel model considers more than 10 unique characteristics related to:

- structure footprint characteristics

- next smaller and larger structure characteristics
- distance to roads
- land cover

The no-parcel model scores each footprint within the unparcelled area, and any structure with a score above the cutoff point is selected as a serviceable location. Figure 10 demonstrates the selection made in an area with no parcels.

Figure 9

3.2.2.3. Visual Verification

Despite the accuracy of the predictions made by BSLPro and Intellistructure, there are still areas of uncertainty, either due to the existence of contradicting data or gaps in the available data. In some cases, two data sources may provide contradictory evidence about whether a structure does or does not exist. This can occur when a tax assessor record indicates there is a home on a developed property with land and improvement value, but there exists no structure footprint. To address these issues, a process of visual user review identifies whether a BSL may exist. This process, known as Visual Verification, employs a managed crowd of trained workers to review satellite and aerial imagery to help identify BSLs. We currently use Visual Verification at scale and in production environments.

To date, approximately 1.5 million records have been sent out to a crowd of trained workers who identify whether the record is a serviceable location based on viewing the source data against visual imagery. To add consistency to the visual review process, every worker is instructed to understand what a serviceable location represents and how to identify them.

At least three reviews of each record are made unless consensus is not reached, in which case additional reviews are triggered until confidence of 75% is reached. If no consensus is reached, then we determine the record to be "Unknown."

After Visual Verification is complete, records are eligible for inclusion in the verified record repository. For every subsequent version of the Fundamental Fabric, we incorporate the information from this repository. Repository information is used to adjust the fit of the models, as well as direct modification of the output when recent, well-matched, verified data are available. Verified data takes precedence over the results of BSLPro and Intellistructure (except in cases of extremely high model confidence).

3.2.3. Adding BSL Attributes

After the Fundamental Fabric is complete, the process moves to attribute development. In these steps, the unit quantity is added, the primary and secondary addresses are selected, the building type code is added, and the bsl flag (bsl_flag) is set.

Most source datasets used for attribute development have been prepared in the Fundamental Fabric. In some cases, the different sources may need to be reconciled and ranked. In other circumstances, sources are ranked against external datasets. By the end of the attribution process, records are identified as Broadband Serviceable Locations.

3.2.3.1. Development of the Building Type Code

Each location record is attributed with a value in the building type code column. The value defines the building as either residential (R), non-residential (B), mixed (X), Group Quarters (G), Community Anchor Institution[26] (C), Enterprise (E), or Other (O). We describe the approach as follows:

- Designation as to residential or non-residential is made based on tax assessor attributes and residential delivery status information provided by address data sources. A mixed location contains both residential and non-residential units.
- Group Quarters and Community Anchor Institutions are identified based on datasets identified in the structure use data sources.
- Enterprise or Special Access locations are location records identified as likely users of non-mass market broadband. These locations were identified based on a special extract of address source data.
- Locations identified as Other fit none of the designations above. They are not categorized as location records and are not found in the Active Fabric output file.

Community Anchor Institutions

CAI locations are identified in our structure-use process. In general, one of two methods is used to determine CAI locations. Table 2 lists all CAI data sources.

[26]The FCC, declined to collect data on non-mass market locations, including healthcare organizations, schools and libraries, government entities and other enterprise customers. See Third Report and Order, paragraph 19

If the CAI is identified using a polygon boundary (such as with prisons, military bases, and universities/colleges), all parcels intersecting the boundary are replaced with the entity boundary, and that polygon is then flagged as CAI.

If the CAI is identified with a point data source, the intersecting footprint is identified as a CAI. In the case where a local government or public safety CAI falls within a Census Designated Place with a population of less than 500,000 people, those points will not be identified as a CAI. Local-government and public safety buildings in places with more than 500,000 people are treated as CAIs.

Group Quarters

A Group Quarters (G.Q.) location is considered a BSLs for the BSLF.[27] These locations are also identified in the structure-use process. As with CAIs, Group Quarters are identified differently based on the data source (see section Implications of Structure Use—Entity Boundaries). If the G.Q. is identified based on a Census 2020 block level count, the G.Q. is identified based on the most likely point in the Census Block to be a G.Q. If Census 2020 indicates there is population in a Group Quarter, but no serviceable location is contained in the Census block, then that Group Quarter count for the block is discarded.

Enterprise or Special Access Locations

Location records requiring special access or enterprise broadband services are identified using methods and data sources similar to those used for the Connect America Cost Model.[28] A special extract of business types (NAICS codes) and employee counts was obtained from a third party data vendor.

A serviceable location was identified as a non-BSL (bsl_flag=FALSE) when at least one business within that serviceable location was in the following NAICS and employee count categories. Figure 11 demonstrates the necessary combinations of employee counts and NAICS codes. Only location records identified as non-residential delivery points are identified in this process.

Segment	Selects
1	NAICS code between 11 and 49 and employee count between 50 to 99
2	NAICS code between 11 and 49 and employee count between 100 to 500
3	NAICS code between 50 and 55 and employee count between 10 to 99
4	NAICS code between 50 and 55 and employee count between 100 to 500
5	NAICS code between 56 and 59 and employee count between 50 to 99
6	NAICS code between 56 and 59 and employee count between 100 to 500
7	NAICS code between 60 and 69 and employee count between 10 to 99
8	NAICS code between 60 and 69 and employee count between 100 to 500

[27] See Third Report and Order, FCC 21-20, paragraph 127.
[28] See Connect America Cost Model (CACM) Model Methodology, https://www.fcc.gov/wcb/CAM%20v.4.2%20Methodology.pdf, Page 20.

9	NAICS code between 70 and 89 and employee count between 10 to 99
10	NAICS code between 70 and 89 and employee count between 100 to 500
11	NAICS code between 90 and 93 and employee count between 10 to 99
12	NAICS code between 90 and 93 and employee count between 100 to 500
13	NAICS code between 11 and 93 and employee count between 500 to 1000000

Figure 11

3.2.3.2. Unit Attribution

We use Census data, tax assessor data, and third-party data to estimate residential housing unit counts. Residential unit counts are assigned to be consistent with Census 2020 housing unit counts 95% of the time measured at the Census Block Group. We estimate non-residential unit counts by taking data from the tax assessor (if provided) along with data from our current third-party vendor.[29]

The residential unit attribution approach is described as follows:

- Gather information regarding units from tax assessor data, Census 2020, and third-party data;
- Remove outlier data by comparing across the sources using an established tolerance;
- Adjust units by targeting structures that are known to be Multi-Tenant Units (MTU) based on their land use, size, or whether the structure was identified by our third-party data source as having multiple units.
- If there are additional units identified by Census 2020, distribute them to known MTUs until those locations are considered full based on their attributes; and
- In cases where Census-provided estimates of units are below what the Fabric has initially estimated, remove units first from multi-unit locations not identified as potential MTUs, then from those that are, if necessary.

Calculated residential and non-residential values are reconciled against entity boundary information. Finally, residential and non-residential unit counts are added together and applied to each location record.

3.2.3.3. Address Attribution

This process assigns every location record a primary address that represents the highest confidence textual address.[30] This is necessary since multiple address data sources are used that may represent the same address at different geographic locations. The assignment process:

- Cleans address records using a known dataset of "bad patterns" that have been identified through multiple versions of development and matching millions of address records to the location Fabric;
- Runs every address record and external address through an address standardization tool. To maximize functionality, we also keep all original address data;
- Develops rankings based on the address ZIP code using geospatial analysis;
- Develops an address quality score;
- Filters out unusable address styles such as post office (P.O), highway carrier (H.C.) boxes, general delivery addresses, and H.C. Boxes;

[29] In the case where our process identifies a home-based business, we count that location as only residential.
[30] There are circumstances where a location record has no address. This can occur in remote areas or areas without commercial delivery services.

- Creates a custom address result code based on the previously created quality score, zip ranking, and address ranking;
- Uses all of the scores and rankings outlined above to identify addresses as primary or identified as not to be used (DNU); and
- Augments the primary address dataset by interpolating addresses based on the geographic location of the Fabric location if no address data source identified a record nearby.

Secondary Addresses

Secondary addresses are assigned to a location record when there is evidence of an assessor-assigned secondary address. A secondary address can be assigned, for example, in circumstances where a building may face different streets or may cross multiple parcels and be assigned different addresses.

The Secondary Address table contains these additional addresses. Each address is assigned a unique record identifier as well as an identifier of the corresponding location_id. An address will be unique in this table by location_id, but multiple secondary addresses may be present for a given location_id.

3.2.3.4. BSL Flag

After identifying location records, additional analysis is required to determine if the serviceable locations should be bsl_flag, TRUE. When a location is a BSL, the bsl_flag is set to TRUE (1).

The bsl_flag is set TRUE for locations utilizing mass-market broadband service. This includes residences, Group Quarters, and non-residential locations not requiring enterprise or special access services.[31] [32]The building_type_code field values of (R,B,X,G) result in a bslf_flag of TRUE. The bsl_flag is set to FALSE when a location record is identified as a certain type of Community Anchor Institution (CAI), a non-mass-market, special access, enterprise location, or other, which uses the building_type_code field values of (C, E, O).

Conflicting Data Source Designations

There are cases when data sources provide contradictory information. In these cases, we have established two rules that override the rules discussed above. First, if a point data source is found to be within an entity boundary representing a college/university, prison, or military reservation, the geographic area boundary determines the nature of the CAI and Group Quarter designation. An example of this is a police station located inside of a college or university boundary. Further, the process uses the area boundary to determine the CAI count.

In the case where a CAI or enterprise customer assignment is made in a location that has a residential land use designation, the process maintains the residential land use designation and overrides the enterprise or CAI designation.

Serviceable Location Record Presentation for CAI and Group Quarters

Defining a serviceable location as a BSL, Group Quarter and/or CAI determines record attribution as well as the number of corresponding records in Fabric output files.

Figure 11 summarizes how serviceable locations are reported based on their designation.

[31] See Third Report and Order, FCC 21-20, paragraphs 126-128.

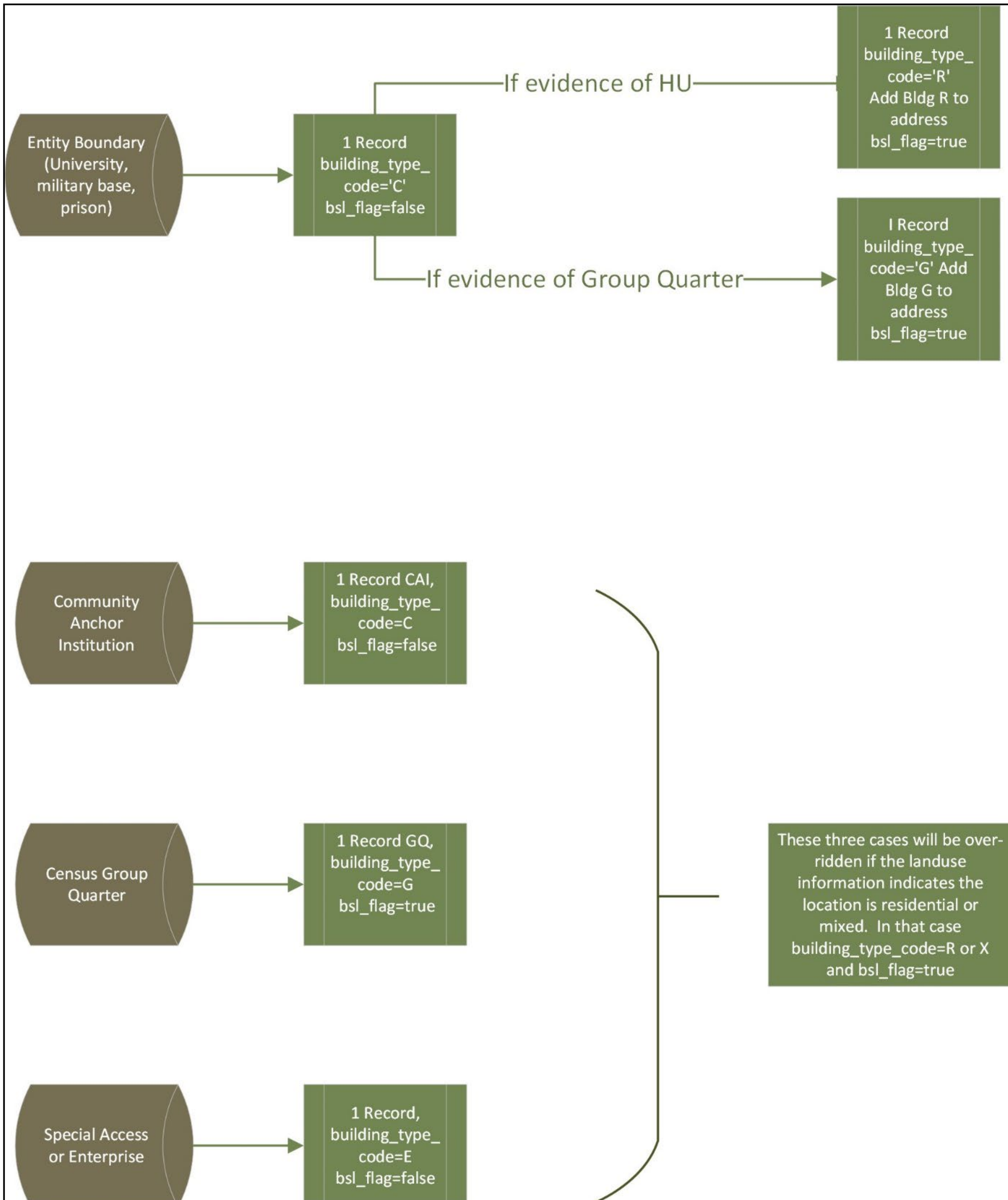

Figure 10

4. Reporting and Output

Fabric output is distributed in a format consistent with the license class of the user.

The Fabric output file is a comma-separated-values (CSV) file containing all Active serviceable locations. The file is unique on location_id and contains attributes such as an address, bsl_flag, and address confidence.

The Secondary Address output file compliments the Fabric output file. It is not unique on location_id, but it will provide additional addresses (if any) at the location_id.

In addition to the Active location files created for licensees, CostQuest provides additional data to the FCC related to other locations and features reviewed as part of the Fabric production process

4.1. Fabric Output Data Dictionary

Column Name	Definition	Unit of Measure	Minimum Scope	Data Type
location_id	Unique ID for the Fabric location. Location_id remains constant when newer evidence indicates the position of the serviceable location or the presence of the serviceable location on a single location parcel is not significantly changed as compared to the prior version.	Unique Key	Fabric Location	int32
address_primary	Postal address (Street number, street name, directional. (street type, city, state and ZIP code excluded)	Indicator (see definition)	Fabric Location	VarChar (190)
City	Postal address city name	Indicator (see definition)	Fabric Location	VarChar (50)
State	Postal address 2-digit state abbreviation. The state is assigned based on information from the assessor data. There can be circumstances where the state differs from the geographic assigned county and Census block.	Indicator (see definition	Fabric Location	VarChar (2)
Zip	Address USPS ZIP Code – A 5-digit code that geographically identifies individual Post Offices or metropolitan area delivery stations	Indicator (see definition)	Fabric Location	VarChar (5)

Column Name	Definition	Unit of Measure	Minimum Scope	Data Type
	associated with every mailing address			
zip_suffix	Address USPS ZIP+4 ®	Indicator (see definition)	Fabric Location	VarChar (4)
unit_count	An estimate of the number of residential, non-residential units within the location. In the case of Community Anchor Institutions and Group Quarters, there is an indication of the presence of CAI or Group Quarters structure.	Count	Fabric Location	int32
bsl_flag	An indication of if the Fabric record is a Broadband Serviceable Location (1) or not (0).	TRUE/FALSE	Fabric Location	Bit
building_type_code	Indicates if the record is considered residential (R), non-residential (B) mixed (X), Group Quarters (G), CAI (C), Enterprise (E) or Other (O)	Indicator (see definition)	Fabric Location	VarChar(1)
land_use_code	A modeled land use designation estimated from assembled county assessor information. - Value may be null. - Value must be one of the following codes, if not null: 1–Residential 2–Land 3–Business 4–Unknown 5–Agriculture 6–Community 7–Industrial 8–Recreation 9–Utility 10–Mixed Use 11–Transportation 12–Water 13–Communications 14–ROW 15–Wireless 0–Other	Enumerated	Fabric Location	int32

Column Name	Definition	Unit of Measure	Minimum Scope	Data Type
address_confidence_code	A code indicating confidence in the association of the textual address with the BSL.	Enumerated	Fabric Location	VarChar(6)
county_geoid	5-digit TIGER 2020 identifier for the county	Federal Information Processing Standard	County	VarChar (5)
block_geoid	15-digit 2020 U.S. Census Bureau FIPS code for the census block.	Federal Information Processing Standard	Census Block	VarChar (15)
h3_9	A reference to the H3 hex cell I.D., level 9.	Indicator (see definition)	Location	VarChar(20)
Latitude	Latitude coordinate of the Fabric location	Decimal degrees (WGS84)	5 digit precision WGS-84	Decimal (11,8)
Longitude	Longitude coordinate of the Fabric location	Decimal degrees (WGS84)	5 digit precision WGS-84	Decimal (11,8)

4.2. Secondary Address Output Data Dictionary

Column Name	Definition	Unit of Measure	Minimum Scope	Data Type
location_id	Unique ID for the Fabric location.	Unique Key	Fabric Location	int32
address_id	Unique ID of Address within Location_ID	Unique Key	Fabric Location	Bigint
primary_secondary	Indicates whether the address is considered primary(P) or secondary(S)	Indicator (see definition)	Fabric Location	VarChar (1)
address	Postal address (Street number, street name, directional, street type with unit (apt, suite, etc.), city, state and zip code excluded)	Indicator (see definition)	Fabric Location	VarChar (190)
city	Postal address city name	Indicator (see definition)	Fabric Location	VarChar (50)
state	Postal address 2-digit state abbreviation. The state is assigned based on information from the assessor data. There can be circumstances where the state differs from the geographic assigned county and Census block.	Indicator (see definition)	Fabric Location	VarChar (2)
zip	Address USPS ZIP Code – A 5-digit code that geographically identifies individual Post Offices or metropolitan area delivery stations associated with every mailing address	Indicator (see definition)	Fabric Location	VarChar (5)
zip_suffix	Address USPS ZIP+4 ®	Indicator (see definition)	Fabric Location	VarChar (4)

