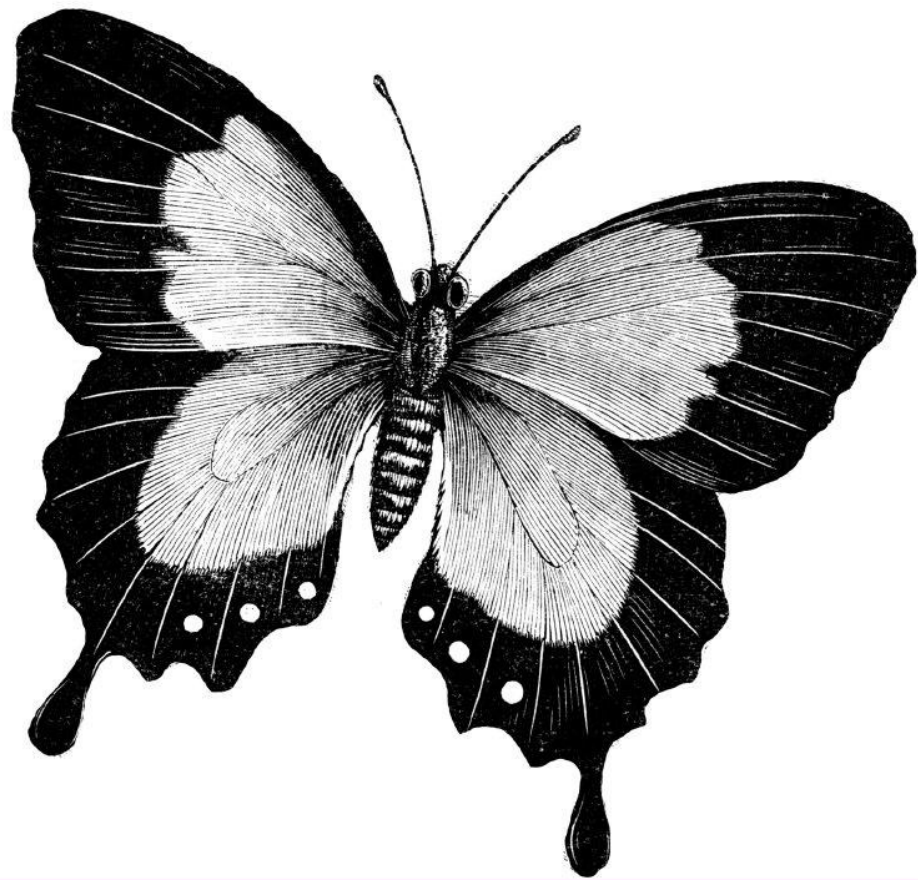


What we know, What we guess, and What we need to find out



Broadband Data Collection

The Missing Manual

O'RELY?

Broadband Commons

Introducing Broadband Data

Introduction

If you’ve found this book, you’ve probably wondered at some point: “who has internet, and how good is it?”

Answering that question is no small feat. For one, there is no definitive report of who’s online and how well their internet works. Where social scientists have a decennial census to probe at and glean new information from, broadband analysts rely on a manifold data landscape to figure out what internet access looks like for a population of users. That can mean a painstaking journey of tracking down raw data, triangulating service from heterogeneous sources, controlling for bias, working through data-hygiene quirks, and more.

This book is intended to be the “missing manual” to broadband data users — how to find the data you need, how to wield it to answer your research questions, common gotchas to look out for, and how to share your work back with the community.

About Broadband Commons

[Broadband Commons](#) is an open collective formed to make broadband data more accessible to the digital equity community. Our output is free and available to all. Anyone is invited to join — [click here](#) to become a Commoner 🙌

Authors

Shaddi Hasan

Shaddi Hasan is an assistant professor of computer science at Virginia Tech.

Robert Martin

Robert is a public technologist and owner of Works Public, a broadband data consultancy. He’s worked with governments from the local to federal level on mapping, software development, and data engineering. Robert is based out of Chicago.

Sascha Meinrath

Sascha Meinrath is the Palmer Chair in Telecommunications at Penn State University, and director of X-Lab. Sascha is a co-founder of MeasurementLab.net and BroadbandMapping.com, and has been at the forefront of national broadband mapping initiatives.

Nick Pappin

Recovering IT person. Credentialed map nerd. Passionate about rural Broadband.

Nick Pappin has worked on almost every aspect of IT from small business desktop support to large scale systems and services supporting an R1 University. Now he works with communities to put together broadband and digital equity plans that will help them know how best to deploy the BEAD funds becoming available to them.

Christine Parker (she/her)

Christine Parker is the Senior GIS Analyst at the Institute for Local Self-Reliance. In her role on the Community Broadband Networks team, Christine is responsible for all map and data-driven projects which include in-depth reports, statistics and maps for articles, and dashboards. Christine also contributes to the team's outreach efforts as the resident mapping expert, and provides guidance on the FCC's National Broadband Map challenge process. Christine holds a Ph.D. in Natural Resources and Environmental Science from the University of Illinois at Champaign-Urbana where she spent a lot of years studying ticks, migrating birds, and wild turkey behavior.

Alexis Schrubbe

Alexis Schrubbe is the Director of the Internet Equity Institute at the University of Chicago. Before joining UChicago, she was responsible for undertaking large scale broadband data collection efforts for the Michigan Moonshot, a program aimed to enhance community-owned broadband in Michigan. Alexis received her PhD from the Technology and Information Policy Institute at the University of Texas where her research straddled broadband access and public policy across multiple geographies in the US and beyond.

Michael Wasser

Acronyms

There is no shortage of them in this space

- ABI:
- ACP: Affordable Connectivity Program
- [BDC](#): Broadband Data Collection
- BEAD: Broadband Equity, Access, and Deployment
- BFM: Broadband Funding Map
- BSL: Broadband Serviceable Location
- [CAF](#): Connect America Fund
- CAI: Community Anchor Institution
- CQA: CostQuest Associates
- FCC: Federal Communications Commission
- FIPS: Federal Information Processing Standards
- FRN: FCC registration number
- ISP: Internet service provider
- LTWF: Licensed terrestrial fixed wireless
- FWA: Fixed wireless access
- MDU: Multi-dwelling unit
- [NBM](#): National Broadband Map
- NTIA: National Telecommunications and Information Administration
- RUS: Rural Utilities Service
- [RDOF](#): Rural Digital Opportunity Fund
- USAC: Universal Services Administrative Company
- UTFW: Unlicensed terrestrial fixed wireless
- WISP: Wireless Internet service provider

Part I: FCC Broadband Data

A question you might have asked getting started with broadband data is — where's the authoritative data? Is there a source of truth that will tell me: "who has internet, and how good is it?"

The answer is: it's complicated. The entity charged with knowing what connectivity looks like nationally is the Federal Communications Commission, or FCC, but their report card is a bit mixed when it comes to recording the reality of where you can *and cannot* access the internet. In the following sections, we'll cover:

- What efforts the FCC has made to measure internet in the U.S., and how well they performed
- What the current process looks like and key things to know about it
- What kinds of insights you can expect to get from FCC data, and where you may want to augment your research with other sources

How Did We Get Here?

While this part of the book focuses on the current iteration of broadband mapping happening at the FCC, it's preceded by a number of earlier broadband mapping efforts.

Form 477 (data collected, data details, data submission, problems)

- Starting in the year 2000, the FCC has collected data about advertised broadband availability for Internet Service Providers (ISPs) via the Form 477.
- Data collected included the [census blocks](#) in which they claim to be able to provide service, the technology used, and the maximum advertised download/upload speeds (i.e., not measures of performance).
- Data was submitted by ISPs twice annually (March & Sep) and those data were later released in June and December, respectively.
- The latest FCC 477 dataset (as of December 2021) is represented in terms of 2020 Census Blocks.
- FCC 477 dataset was particularly problematic because if a provider claimed to be able to serve at least one location within a census block, the entire block was considered served.
 - Impacts on measures of realistic landscape of broadband availability
 - Impacts on measures of broadband competition
 - More info detailed here (dated, but still relevant):
<https://communitynets.org/reports/profiles-monopoly-big-cable-and-telecom>
- Another issue with this dataset is that in addition to over- representing where an ISP can provide service, the measures of speed represent the highest speed tier the ISP *may* offer in that area, and not what a customer may actually be able to receive. For instance, a customer could end up calling the provider or visiting their website only to find out that

the ISP offers plan(s) with lower speeds than what was reported for the census block that their address falls within.

- Criticism
 - [salty tweets and congressional hearings etc here]

Broadband DATA Act

TODO talk about this as the catalyst of BDC

New Data Dawns

Clearly, the status quo of Form 477 was not working, with all of the data-quality and usability issues that were flagged by industry, communities, and government insiders alike. What came next was a concerted effort to rethink how the FCC collects broadband data.

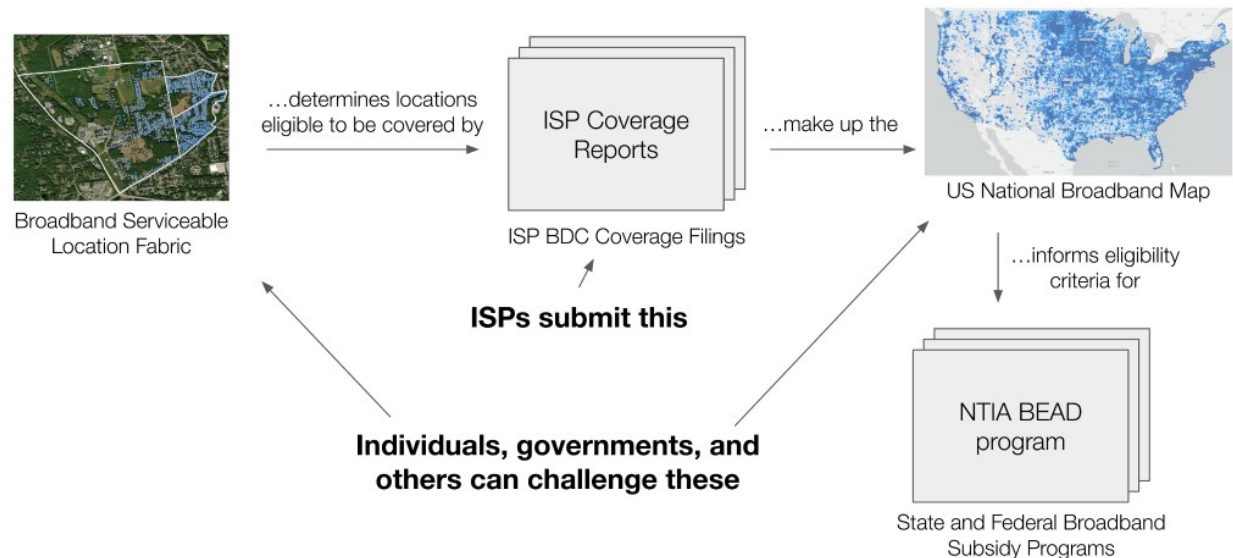
What new processes, products, and concepts were introduced with this large-scale pivot in how mapping is performed at the FCC? In the following sections we'll introduce you to the current state of mapping affairs at the agency.

Broadband Data Collection Program

What is it?

The Broadband Data Collection program (or “BDC” for short) is a new process implemented by the FCC in fall of 2021 to solicit broadband availability from ISPs, as well as challenges to that claimed availability. It represents a part of the agency's broader efforts to overhaul broadband availability data, both in terms of how finely it's measured and how it's made accessible to the public.

How the BDC Process Fits Together



Broadband Serviceable Location Fabric

- What is it?
 - A Very Expensive Map of U.S. Homes and Businesses™
 - A map of buildings that should, in theory, be able to subscribe to consumer-grade high-speed internet
 - In response to the problems with the Form 477 problems Congress passed the Broadband Deployment Accuracy Technology Availability Act (or Broadband DATA, for short).
 - The FCC was charged with creating a national location database of broadband serviceable locations (BSLs), now known as the Broadband Serviceable Location Fabric (or “the Fabric” for short)
 - In addition to creation of the Fabric, the BEAD Act also requires providers to report service availability at the individual BSL (TODO: References below) level.
- What it is not?
 - The fabric does not contain information about internet service, such as which ISPs serve a location, what speeds they offer, etc. It is simply a map with latitude-longitude coordinates for locations where a home or business should be able to get broadband.
 - The fabric specifically excludes any location that is not a residence or business. The rule of thumb is: would this location normally subscribe to mass-market internet?
 - This is missing data that is not end-user servicing. Examples would be:
 - Middle mile fiber lines
 - Backhaul fiber lines

- Internet Exchange Points
 - Data Centers - may be on the map but not mass market
 - Broadband availability data
 - Service offerings, only **max advertised**
- It is also not the web app found here: [Home | FCC National Broadband Map](#)
 - This was created by Emprata. ([FCC Broadband Data Collection contract for Data Architecture and Design Services awarded to Emprata – Emprata](#)) (TODO: Further reference needed)
- How was it created?
 - Ostensibly, parcel data + ...
 - YouTube video queued to the right timecode
[<https://youtu.be/vb7vIORyH54?t=620>]
 -
- How to get it? (Licensing - Do you need one, how to get it, and what it means)
 - Do you need one? - Shaddi has a great article about this, [You Probably Don't Need the BSL Fabric | by Shaddi Hasan | SPIN@VT | May, 2023 | Medium](#)
 - How to get it? - [FCC Broadband Serviceable Location Fabric - CostQuest](#)
 - Fabric data dictionary - [[link](#)]
 - Active, BSL File [[COA-FCC Data Dictionary-Client V2 Active - BSL.xlsx \(cqafabric.s3.amazonaws.com\)](#)]
 - This describes the data about BSL's that is not public ([Note below](#))
 - Active, No BSL File [[COA-FCC Data Dictionary-Client V2 Active - NonBSL.xlsx \(cqafabric.s3.amazonaws.com\)](#)]
 - This is the same data as the Active BSL File but about locations that are flagged as not BSL's (Schools, Libraries, etc)
 - Secondary File [[COA-FCC Data Dictionary-Client V2 Secondary.xlsx \(cqafabric.s3.amazonaws.com\)](#)]
 - This is a secondary address file. It is unclear if these are addresses that have been submitted as corrections and will be secondary forever or will eventually be listed as the primary address in the Active BSL data.
- Summarizing Fabric Challenge Data
 - Challenge outcome data [[link](#)]
 - Some data is not geocoded in the file above. There is a tool built by the Broadband Commons team and available here. [[bdc-data-tools/challenge-lookup at main · broadband-commons/bdc-data-tools \(github.com\)](#)]

National Broadband Map

What it is. The National Broadband Map is the FCC's dataset capturing where broadband service is available on a *per-location* basis.

- What it could be...

Development

- Policy summary that led to this
- Drama that ensued and the players involved (e.g., CQA, FCC)

Where does the data come from?

Everything about the service availability reported in the National Broadband Map starts with data self-reported by ISPs.

ISPs are required to submit data every six months outlining the *list of locations* at which they provide service, the *highest speed service plan* they offer at each location, as well as other metadata about their service offerings (such as the type of technology used to serve that location). In addition, ISPs submit information about the number of subscribers they have in each census block they serve, similar to what was previously submitted via the Form 477 process. The subscriber data seems to be used as a quality check on the service availability data, which is the main focus of the BDC process.

There is no single methodology all ISPs use to submit this data. The general rule is that an ISP can claim a location as served if they could, if requested, provide service at the advertised speed to a location within ten days of a request. There's room for interpretation here: two well-intentioned ISPs could make different reasonable assumptions in translating this guidance into a list of locations they serve, resulting in different reported service availability.

The service availability challenge process. When an ISP reports they provide service at a given speed to a location, there's no verification that the offered speeds in a location are actually correct (e.g., via network measurements). To account for this, the BDC process also allows for individuals, governments, ISPs, and others to *challenge* the service an ISP reports to a location.

- Bulk vs individual challenges
- Challenge adjudication process: challenger and ISP hash it out, if they can't agree, the FCC gets the final say. Each step can take 90 (?) days.

How often is this data updated?

The public availability data sets appear to be updated every other Thursday however the data is as of the previous day and notated as such. This appears to occur as filers update their availability against the fabric. Such as in the case of an ISP accidentally stating they have services available in a neighborhood when it is on their build out list. Additionally we have seen the data get updated after the minor update window. This has usually happened because the data exports were broken.

As a tool to assist users in ensuring they are talking about we propose and have adopted a date based semantic versioning style based off of date. For the major version we use the fabric date of the base map. The minor version is the availability release date. Finally, the bug version is used to connote any additional out of band updates. This may be related to an issue where a file didn't generate correctly or other data cleanliness issues.

In practice this would look like 20020630.20230426.2 for our current release. The fabric is from 2023-06-03, the availability is from 2023-04-26 and there was a problem with that release which led to the bugfix version being 2 due to an issue where fiber data from Utah was blank.

Can we just write a quick calendar of when we should be up in arms about things. Like I know every December, January, June, and July we need to be putting dots on things. Then in February, March, August, and September We need to worry about what the dots say. And November and May are for freaking out about how messed up it all is. This leaves April and October for vacation.

API

The National Broadband Map pulls in data on the fly from an undocumented REST API. This section documents it, in the name of science.

Note: while the National Broadband Map API is “public” in the sense that anyone can access it via an unauthenticated HTTP client on their machine (aka a web browser), we do not know what the FCC's official position is on accessing the following endpoints outside of surfing the National Broadband Map. As a general rule, if the data contains addresses or latitude-longitude coordinates, it's property of CostQuest, and scraping their data doesn't appear to be condoned. Talk to a lawyer, be respectful, and use at your own discretion.

The base URL for the API is: `https://broadbandmap.fcc.gov/nbm/map/api`

- `GET /fabric/detail/{PROCESS ID}/{LOCATION ID}`
 - Returns details for a location

Fixed Availability Data

For each broadband serviceable location included in the CostQuest Fabric, there is corresponding *availability data* that describes the “make and model” of the service an ISP claims to be able to provide. These data are not to be taken as indications of performance, and should always be regarded with a healthy dose of skepticism because ISP’s are known to make claims of service availability beyond where they can reasonably offer service, both in terms of physical location and the technological capacity of the technology used. An example might be a wireless Internet service provider (WISP) claiming to offer symmetrical service (100/100 Mbps) throughout their entire network footprint - this is just not realistic. In the sections below, we’ll describe data that ISPs submit that are not public, and those that are public.

Not Public

- Location data - lat/long/address/MDU [these are the propriety CQA data]
- Subscriber data [contains PPI, but seems like ISPs also consider this proprietary?] - ideally these data would be aggregated to census block or other geography and shared publicly

Public Facing Data

Availability Data Methodology

Have you ever wondered how they arrived at those collections of polygons and addresses they’re uploading to the BDC portal? Wonder no longer — it’s open data 🙌

If you navigate to the “Data by Provider” download page and start typing in the name of a provider, you’ll see “Supporting Data” to the right. Those files, broken out by technology, contain methodology abstracts for how ISP determined served locations. All of the following describe how fiber-to-the-premises location were calculated, with the exceptions of SpaceX and Viasat (both satellite).

What is an FCC FRN?

An FRN, or FCC registration number, is a 10-digit number that is assigned to a business or individual registering with the FCC. This unique FRN is used to identify the registrant's business dealings with the FCC. All ISP’s are required to file broadband availability data with the FCC, and to do so they must obtain an FRN to access the Broadband Data Collection system.

A fun twist is that some companies have more than one FRN! If you are querying data for specific companies, this is an important fact to be aware of. Companies may have different FRN’s for different parts of the business. For example, a telephone cooperative may have an FRN for the telephone side of the business and a different FRN for their broadband subsidiary.

If you are unsure what the FRN number is for a particular company, you can use the [FCC's FRN search tool](#). The downside to using this tool is that in addition to companies having more than one FRN for different parts of the business, you may also discover additional FRN's listed for the same company that are registered by different people, in different years, and sometimes at different addresses. Another caveat associated with this tool is that any company that conducts any business with the FCC must have an FRN, so there are many companies in the database that do not provide Internet service.

A much better resource for finding company-specific information is the [FCC's BDC Provider ID Table of Service Providers](#). This table includes the local provider name, holding company name, provider FRN, holding company number, and the company's operation status (ILEC¹/non-ILEC).

Holding Company Data

Any provider's holding company can be tracked down in [this dataset](#). You might be wondering what a holding company is and why you should care about it. A [holding company](#) tends to oversee managerial decisions, but typically plays no role in day-to-day operations of a business. We care about holding companies because it helps us better understand the state of competition in the ISP marketplace. For example, in some geography you might have 4 different providers offering service, but 2 of them are actually owned by the same holding company. So for those 2 providers, profits are going to the same holding company who is also likely making decisions to update infrastructure (or not) and determine pricing plans that influence our day-to-day interactions with the service.

Maximum Advertised Speeds

Each record of data in the BDC includes the maximum advertised download and upload speeds for a given BSL. These speeds are reported in Mbps (megabits per second) and do not necessarily indicate the speed that the end-user subscribes to and/or receives. Rather these speeds are what the ISP *claims* they can provide at the locations.

Latency

This field indicates whether the reported service is low latency (value = 1) or not (value = 0). Low latency is defined as having round-trip latency measure of less than or equal to 100 ms. This measure was not reported via Form 477.

¹ ILEC refers to an Incumbent Local Exchange Carrier. [This page](#) provides a deeper dive into ILECs and Competitive Local Exchange Carriers (CLECs).

Technology

This field indicates the type of physical infrastructure that is used to deliver service to a given BSL.

- 10 – Copper Wire
- 40 – Coaxial Cable / HFC
- 50 – Optical Carrier / Fiber to the Premises
- 60 – Geostationary Satellite
- 61 – Non-geostationary Satellite
- 70 – Unlicensed Terrestrial Fixed Wireless
- 71 – Licensed Terrestrial Fixed Wireless
- 72 – Licensed-by-Rule Terrestrial Fixed Wireless
- 0 – Other

Be aware that if you are comparing BDC data with data from Form 477 the technology codes are slightly different.

ILEC Study Area Boundaries

- The FCC began to collect spatial data that describes where ILEC's offer telecommunications service. These areas were "frozen" in 1984 to prevent the development of a telephone provider monopoly. Details of the order that initiated this data collection and the wiggle room built in for monopoly-providers can be [found here](#).

Broadband Funding Spent on Infrastructure

BTOP - [Broadband Technology Opportunities Program](#) (BTOP)

CAF - [Connect America Fund](#) [FCC]

CAF II - [Connect America Fund Phase II Auction](#) [FCC - data under 'Data' and "Results" tab] or [Auction 903 Results](#)

(FCC Open Data - spatial dataset; not up to date)

RDOF - [Rural Digital Opportunity Fund](#) [FCC]

Both the Connect American Fund and Rural Digital Opportunity Fund were intended to provide support for building infrastructure that would permit Internet access at unserved locations².

Location-level data for the programs above can be [found here](#).

[ReConnect](#) [USDA]

[Technology Opportunities Program](#) - Physically archived, but unsure if digitally available somewhere

² The definition of served for CAF was 10/1 Mbps and 25/3 Mbps for RDOF. (check this)

BFM - [Broadband Funding Map](#) > This map illustrates locations for which an ISP has received funding to build infrastructure to ensure the location is served. A major caveat to this map relates to how Internet availability and infrastructure grants were previously mapped, at the census block level.

When asked whether BSLs were known to have received funding or whether they were simply classified as funded because of their census-block location, we received the following response from the Broadband Data Collection:

“Every BSL is already associated with a Census Block when it was created for the Fabric files sent to Providers in order to be used to submit their data. BSLs are spatially associated with the blocks as any address or geographic coordinate would be.

The funding data that supports broadband buildout shown on the map are provided by Federal agencies and therefore we can not speak to their methods or implementation.”

In order to determine how BSLs were related to federal funding remains unclear.

The data are available for download, in a process similar to the BDC. However, not all program data are available in the same format. Some are available as polygons in geopackages, some are available as location lists in csvs, and a few have both of those data types available.

The location lists are further complicated by the fact that some locations include a location id, but no census block or other typical information that could permit analysis or mapping. The other locations have latitude and longitude but do not have a location id.

Format aside, the data has some additional caveats that folks should consider.

- Not all records include a provider id - so if you're looking for provider-specific funding, provider id may get you part of the way, but you may need to search by brand name.
- Some records will include multiple technologies (separated by ;) - this means you won't necessarily be able to visualize or analyze by single technology types for all programs.

Broadband Funding Map Data

Broadband Funding Spent on Maps

\$350 million re: 2010 mapping/original broadband map, BTOP

\$45 million FCC Broadband Fabric

\$??? CPF/SLFRF

\$48 million NTIA Broadband Fabric Access?

??? State funding

??? RUS/ABI (2017-2021 era)

??? NTIA Measures of Broadband Need
??? super-secret “internal” map (apparently NTIA dev’ed it, but wouldn’t share w/ other agencies for BEAD)
\$??? TOP

Analyzing FCC Data

Broadband Commons has done a lot of work to help folks be able to better access and work with the Broadband Data Collection. Those resources can be found at our GitHub page.

So you don’t want to individually download 200ish files?

We have a tools for that!

<https://github.com/bbcommons/bbkit>

Need hexagons for mapping or analyzing data?

You can download individual states or the whole country. <https://github.com/bbcommons/data>

Spatial Analyses

- Spatial analysis of BEAD-defined served/unserved using standardized geography (e.g., state, zipcode, county, etc...)

Version Changes

- Evaluate changes between map versions (major (6 month) and minor (2 week) updates)

Process issues

BSL Updates to the fabric (Unit changes, Address updates, etc.)

How to find help analyzing broadband data

What kind of people, their skills, their rates, how much time it takes

Part II: Additional Broadband Data

How to Not Use the Fabric

Author: Robert (and calling all collaborators)

Synopsis: You don't need the CostQuest fabric to do your work. Roll your own with open data!

M-Lab

Introduction

What can I use M-Lab data for?

Ookla

Netrics

Glossary

- **BSL Fabric:** A database of points that underlies the National Broadband Map. Each point roughly represents a building that could receive a high-speed Internet connection. This database is maintained by CostQuest Associates, under contract to the FCC.
- **FCC National Broadband Map:** This is the map that shows service availability, based on both the BSL Fabric as well as data provided via the BDC process.
- **Broadband Data Collection:** The FCC program that defines the processes by which data is collected to create the National Broadband Map. The BDC includes both the process by which data is collected from Internet service providers every six months, as well as the service availability challenge process, by which individuals, governments, ISPs, and other organizations can challenge the reported availability of broadband on the National Broadband Map.
- **Form 477:** The name of the form used for collecting service availability data from ISPs prior to the introduction of the BDC process. Sometimes used interchangeably with the dataset represented by the previous version of National Broadband Map, which used Form 477 as its data source.

