

# **DATA DEVELOPMENT & VALIDATION METHODOLOGIES WHITE PAPER**



## **Commonwealth of Kentucky State Broadband Initiative (SBI) Broadband Mapping Project**



**NTIA Data Submittal  
September 30, 2011**

**Baker**

## Table of Contents

<b>Introduction .....</b>	<b>4</b>
<b>Broadband Provider Outreach .....</b>	<b>4</b>
Data Submission Guidelines .....	4
Kentucky Broadband Providers Website .....	4
Outreach Delivery Vehicles .....	4
Broadband Outreach Tracker Application .....	5
<b>Provider Submittal Validation .....</b>	<b>6</b>
Validation Checklist .....	6
Data Usability Determination .....	7
<b>SBI Data Development .....</b>	<b>7</b>
2000 to 2010 Census Data Translation .....	7
Spatial Data .....	7
Address Data Geocoding .....	7
Census Block Aggregation .....	7
Road Segment Aggregation .....	8
Overview Data Aggregation .....	8
Polygonal Boundary Aggregation/Integration .....	8
Wireline Provider .....	8
Wireless Provider .....	9
Middle/Last Mile Data Integration .....	9
Community Anchor Institution Integration .....	9
Provider CAIs .....	9
Commonwealth CAIs .....	9
<b>Provider Validation .....</b>	<b>9</b>
Types of Provider Maps .....	9
Outreach Maps .....	9
Initial Verification Maps .....	10
Detailed Verification Maps .....	10
Revised Maps .....	10
<b>Data Validation .....</b>	<b>11</b>
Validation Data Set Collection and Development .....	11
Provider Data Validation Process .....	16
<b>Validation and Confidence Level Reporting .....</b>	<b>16</b>
Low Confidence Provider Feedback .....	18
<b>Changes and Corrections Documentation .....</b>	<b>18</b>
<b>Future Validation .....</b>	<b>18</b>

**Table of Figures**

Figure 1 Broadband Outreach Tracker .....6

Figure 2 Provider Map ..... 11

Figure 3 Field Verification Sampling Locations..... 13

Figure 4 Statistical Evaluation and Assessment System (SEAS)..... 17

**Appendices**

**Appendix A: Commonwealth of Kentucky SBDD Process Workflow.....19**

**Appendix B: State Broadband Data Validation Workflow.....20**

**Appendix C: 2000 to 2010 Census Data Translation Workflow.....21**

## Introduction

The following sections of this document provide an overview of the process used for the SBI Broadband Mapping data development for the Commonwealth of Kentucky. The following narrative is depicted in Appendix A, Commonwealth of Kentucky SBI Process Workflow, and Appendix B, State Broadband Data Validation Workflow, included at the end of this document.

## Broadband Provider Outreach

The following outreach procedure provides the framework for communicating with Broadband Service Providers (Providers). The primary goals of the outreach approach documented herein are to:

- Promote Provider understanding and acceptance of the Broadband Mapping process, results and benefits
- Clarify NTIA Broadband Mapping requirements
- Facilitate data confidentiality agreements as required
- Minimize the submittal of invalid data
- Enhance provider understanding of the semi-annual update process
- Work with Providers to evaluate submittal options to facilitate data submittals

## Data Submission Guidelines

Guidelines for the Providers submission of of Broadband Mapping Data are documented in the “Data Submission Guidelines”. These Guidelines define technical requirements, submission, specifications, and coordination and documentation activities.

## Kentucky Broadband Providers Website

A URL was deployed (<http://www.bakergis.com/kyBroadbandProvider/>) to communicate and distribute NTIA NOFA requirements to providers along with outreach and data submittal materials including:

- NTIA NOFA and subsequent clarification
- Outreach letters to providers
- Non-Disclosure Agreement
- Quick Start Guides
- Data Submission Guidelines
- Data Transmittal Letter
- Broadband Data Submittal Templates
- TIGER Data
- Data Submittal Assistance Contact Information

## Outreach Delivery Vehicles

- A State Broadband Mapping Initiative Call for Data letter from the Kentucky Commonwealth Office of Technology (COT) was emailed to all Broadband Service Providers in the Commonwealth. This initial provider contact letter described the program and the role of Michael Baker Jr., Inc. (Baker) acting on behalf of the COT for Broadband Data Collection and Mapping.

- Baker distributed a follow-up letter to all Providers describing the data submittal requirements and material and help available to aid with the data submittals.
- Submittal assistance was provided to providers that needed help with data submittals.
- Presentations were conducted with various broadband provider associations to present the data submittal requirements and answer questions.
- Email communication and electronic transfer of data was encouraged to facilitate a faster delivery of data and information.
- A URL was deployed and promoted to distribute outreach material and information concerning the Broadband Mapping Project.
- A secure FTP URL was provided for submittal of broadband data by providers.

### Broadband Outreach Tracker Application

The Tracker application (Figure 1) was utilized to collect all correspondence with Providers and feedback on the effectiveness of the outreach activities by tracking items such as:

- The number and content of incoming e-mails and letters submitted from the Providers
- The number and source of comments, questions, and suggestions made by Providers
- The number and source of comments, questions, and suggestions made by attendees at Provider meetings and conference calls
- Provider contact information and data submittal status.

The screenshot displays a web-based form titled "Broadband Outreach Tracker" with three main sections: Business, Legal, and Technical.

**Business Section:**

- Delivery Type: [Dropdown]
- Agreed to Participate: [Dropdown]
- Date to be Delivered: [Calendar icon]
- Comments: [Text area]
- Date Last Updated: [Text field]
- Last Updated By: [Text field]

**Legal Section:**

- Date NDA Received: [Calendar icon]
- Returned to Provider: [Calendar icon]
- Screened for Changes: [Dropdown]
- NDA Executed & Returned: [Calendar icon]
- Date Last Updated: [Text field]
- Last Updated By: [Text field]

**Technical Section:**

	Date Data Received	Data Complete	Date First Screened	Data Accepted	Broadband Data Accepted
D1	[Calendar icon]	[Dropdown]	[Calendar icon]	[Dropdown]	[Calendar icon]
D2	[Calendar icon]	[Dropdown]	[Calendar icon]	[Dropdown]	[Calendar icon]
D3	[Calendar icon]	[Dropdown]	[Calendar icon]	[Dropdown]	[Calendar icon]
D4	[Calendar icon]	[Dropdown]	[Calendar icon]	[Dropdown]	[Calendar icon]

Below the table:

- FTP User: [Dropdown]
- FTP Date: [Calendar icon]
- Date Last Updated: [Text field]
- Last Updated By: [Text field]

Figure 1 Broadband Outreach Tracker

## Provider Submittal Validation

When a data submittal is received from a broadband service provider it is updated in the Broadband Outreach Tracker and run through an initial validation process to assure that it meets the submittal guidelines.

### Validation Checklist

The following items are part of this initial data validation process:

- Verify the provider Transmittal Letter is complete and matches submitted data
- Verify the file naming conventions
- Verify each file is machine readable
- Verify data is in the correct GIS or Tabular format/file type
- Verify there are no duplicate records
- Verify each field is populated and no empty or NULL values are present for mandatory fields
- Verify all ID (record number points) are unique within the submittal
- Verify all attribute data is formatted according to the submittal guidelines
- Verify topology for all geospatial submissions
- Verify Metadata for all submissions
- Verify the required contact information is included
- Verify adherence to Data Submittal Guidelines (see <http://www.bakergis.com/kyBroadbandProvider/> to access Data Submittal Guidelines)

#### **Broadband Service Availability** (at least one)

- Individual Street Addresses (Sec 3.1 & 4.1)
- Census Blocks < 2 sq mi (3.3 & 4.3)
- Street Segments for Census Blocks > 2 sq mi (3.2 & 4.2)
- Service Overview (Sec 3.4 & 4.4)
- Polygonal Boundary Area(s) (Sec 3.8 & 4.8)

**Middle-mile Points** (Sec 3.5 & 4.5)

**Community Anchor Institutions** (Sec 3.7 & 4.7)

**Last Mile Connection Points** (Sec 3.6 & 4.6)

**WISP Antennas** (Sec 4.9)

## Data Usability Determination

The validation results are evaluated by the outreach and aggregation persons to determine the usability of the data. If the data meets the submission specifications, it is forwarded on for data aggregation. If it is determined to be unusable, it is returned to the Broadband Service Provider for resolution. If the data can be manipulated to get it into a usable format, it is manipulated as required, and then forwarded on for data aggregation.

## SBI Data Development

Data from the Broadband Service Providers may be submitted in various formats as defined in the Data Submittal Guidelines, or in some cases unspecified formats may be accepted to help facilitate provider participation. Depending on the format of the submitted data, it is processed through one of the following processes to upgrade it to the NTIA SDBB data standards.

## 2000 to 2010 Census Data Translation

Many providers indicated there were no changes to their previous data submission that was compiled to 2000 census information and they did not have the capacity to upgrade their data to 2010 census information. Therefore, the Broadband Mapping team has translated the April 2011 data for these providers using the workflow shown in Appendix C.

## Spatial Data

After validation and any required manipulation of any spatial data submitted by the Broadband Service Providers, it is georeferenced and simply loaded into the appropriate NTIA geodatabase feature class.

## Address Data Geocoding

If not already in the standard address point template, the provider tabular address data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. ArcGIS geocoding tools are then utilized geospatially locate the address points for the tabular records. Interactive address rematching is performed against two additional street centerline datasets as needed to increase geocoding matching results. The NTIA deliverable is the geocoded address point geodatabase table. The geocoded address points are also subsequently aggregated to the census block or road segment feature class for public web map display.

## Census Block Aggregation

If not already in the standard census block template, the provider tabular census block data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. The Provider tabular census block records are then joined to the geodatabase 2010 U.S. Census Block. This join is performed as many times as necessary for multiple Trans Tech values for each Provider/Census Block combination. The NTIA deliverable is the census block geodatabase table.

If the list of census blocks contains blocks > 2 sq. miles then these blocks are used to select all the 2010 U.S. Census TIGER centerlines that intersect those blocks. The Census Block record data is aggregated to each Road Segment within the Census Block. This process is performed as many times as necessary for multiple Trans Tech values for each Provider/Census Block combination.

### Road Segment Aggregation

If not already in the standard road segment template, the provider road segment data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. If the provider submittal included graphic centerline segments, these are migrated into the delivery geodatabase along with the linked attribute records. If the provider submittal was tabular road segment records only, they are then joined to the geodatabase 2010 U.S. Census TIGER centerline feature class. This join is performed as many times as necessary for multiple Trans Tech values for each Provider/Road Segment combination. The NTIA deliverable is the road segment geodatabase table.

If the provider road segment data lie within census blocks  $\leq$  2 sq. miles then the road segment data is aggregated to the census block. This process is performed as many times as necessary for multiple Trans Tech values for each Provider/Road Segment combination. The NTIA deliverable is the road segment geodatabase table.

### Overview Data Aggregation

Provider Service Availability Areas submitted for entire county areas are loaded into the NTIA geodatabase Overview table. If not already in the standard template, the provider data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. The Provider Overview records are then joined to the geodatabase 2010 U.S. Census County feature class. This join is performed as many times as necessary for multiple Trans Tech values for each Provider/County Area combination.

### Polygonal Boundary Aggregation/Integration

Broadband Service Providers submitting polygonal service area data is handled in two ways. Wireline Provider data is aggregated to the census block feature class for areas where census blocks  $\leq$  2 sq. mi., or road segment feature class for areas where census blocks > 2 sq. mi. Wireless Provider Service Availability Areas submitted by polygonal area are simply loaded into the NTIA geodatabase Poly\_Bndry feature class.

#### Wireline Provider

The polygonal data is georeferenced and loaded into the Poly\_Bndry feature class. The polygon is then attributed, manually if necessary. Depending on the area, census blocks  $<$  or  $\Rightarrow$  2 sq. mi., a selection set of either census blocks or road segments that intersect the polygon boundary is created. The attributed polygon boundary is then joined with census blocks or road segments table to attribute accordingly. This join is performed as many times as necessary for multiple Trans Tech values for each Provider/County Area combination. The NTIA deliverable is the census block or road segment geodatabase table.



### Wireless Provider

The polygonal data is georeferenced and loaded into the Poly\_Bndry feature class. The polygon is then attributed, manually if necessary. Multiple Poly\_Bndry records are created for multiple Trans Tech values for each Provider. The NTIA deliverable is the polygon boundary geodatabase table.

### Middle/Last Mile Data Integration

If not already in the standard template, the data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. The point features are geo-located utilizing the lat/long information provided. The NTIA deliverable is the middle or last mile geodatabase table.

### Community Anchor Institution Integration

Broadband Service Providers provided some Community Anchor Institution (CAI) data with the data submittals. But the majority of the data was collected from existing GIS Layers maintained by the COT on their KYGEONET public website. Some of the data was collected by outreaching to CAIs through state agencies and their contacts, and having CAIs complete an online survey at [http://www.bakerbb.com/ky\\_institution\\_survey/](http://www.bakerbb.com/ky_institution_survey/).

### Provider CAIs

If not already in the standard template, the data is first loaded into that template. The data is then exported to a geodatabase table using the ArcGIS Conversion Tools. The point features are geo-located utilizing the lat/long information provided. Address data is used to geocode locations only when Lat/Long data is not provided.

### Commonwealth CAIs

CAI shapefiles were downloaded from the KYGEONET website. The shapefiles were then exported to the NTIA geodatabase CAI feature class. Various sources for obtaining broadband information for the CAIs were utilized. Various state agencies provided some of the information, i.e.; Council on Post Secondary Education (CPE) provided tabular broadband information for schools and libraries and COT provided tabular broadband information for health departments. A CAI data survey website was also deployed and the URL distributed by various state agencies to the CAI contacts. Data from all of these sources were then aggregated into the CAI geodatabase table for the NTIA deliverable.

### Provider Validation

After data development, service availability maps are generated and submitted to the providers to validate their mapping results. This provides a “sign off” on the interpretation of the submitted data and extends the outreach efforts by providing a visual representation of the data to be delivered to the State and the NTIA.

### Types of Provider Maps

Provider maps (Figure 2) generally consist of the following types.

#### Outreach Maps

Often time’s providers will send data which does not contain all the information needed for a NTIA compliant dataset. In such cases, as an aid to the outreach communication, it may be necessary to produce a map to help the provider locate their service area or verify data they have provided. These maps may take many forms, but generally are of two types:

- General Location Maps – these maps are often produced when the provider does not have a list of address or other standard submittal data and needs help defining their service area. A typical map will show counties, major roads, and towns of the general area the provider has stated as their service area. The intent of the map is to give the provider a way to markup or delineate their service area. If a provider has not provided required attribute information such as Technology of Transmission, Speed Data, etc. then it may be necessary to add a visual clue to this data like an information stamp on the map that they can easily fill out. If the provider sends the map back with a service area boundary, this can then be digitized and sent back to the provider for verification.
- Verification of Provider Supplied Boundaries – these maps are produced when the provider has sent service area boundary information which is confusing or otherwise unclear. Often these are produced when providers send CAD maps, hand drawn maps that need digitization, or lists of zip codes or counties served. A typical map will place the interpreted boundary over a location map so the provider can verify the service area. As with the General Location Map, information stamps or other visual clues may be placed on the map.

### Initial Verification Maps

Once the provider data has been processed and the census block and road segment feature classes created, an Initial Verification Map is produced to give the provider a visual representation of their service area by census block. These maps enable the provider to verify their service area and make changes if necessary. Initial Verification Maps are produced using a set of standards and produced at the highest resolution necessary to convey the map information to the provider. Initial Verification Maps are also produced for Wireless Polygon areas.

### Detailed Verification Maps

Providers who have questions about their service areas may request additional information to help clarify issues. In these cases it may be necessary to create a Detailed Verification Map to highlight the areas in question. Detailed Verification Maps provide the same information as Initial Verification Maps only at a higher resolution. Several maps may be needed to accurately portray an area in question.

### Revised Maps

Revised maps take two forms:

- Initial or Detailed Verification Maps which have been annotated or marked-up by the provider
- Outreach produced Initial or Detailed Verification Maps incorporating provider changes



Road Segment Coverage as depicted on broadband maps is defined as a 500 foot buffer around existing roads in census blocks greater than 2 square miles in area. Unnamed and other lesser roads may not be shown on the maps. Absence of road features does not necessarily indicate broadband service is unavailable.

— Road Segments: Census > 2 sq mi  
 500 ft Road Segment Buffer  
 Census < 2 sq mi

### Figure 2 Provider Map

A critical component of the project is the validation of the data submitted by the broadband service providers. Data from various sources, as described in more detail in the following sections, is utilized to develop a level of confidence in the data received from the broadband providers.

This validation process employs data sets developed or acquired from different sources as described in the following sections.

Provider Feedback Loop: Maps of completed Provider service areas and data are furnished back to the Providers for confirmation of the processed/aggregated information. Feedback is integrated into the each Provider's dataset.

Strategic Networks Group (SNG) Wireline Market Intelligence Data: Data is extracted from internal and commercial databases defining geographic service areas of telephone and cable companies and locations of central office (CO) switches and areas upgraded with fiber. The geographic areas are overlaid with Census demographic data on housing unit counts and density. The areas are then modified based on standard business practices for conducting service build-out and offering broadband service relative to housing density and other variables, such as distance from CO and other infrastructure elements, type of cable franchise (e.g., Census Place vs. Unincorporated County) This represents the first pass conservative estimate of coverage.

The above methods and data sources are supplemented by other data sources and methodologies, including: 1) connectivity data points acquired from InfoUSA that include ISP and type of connection (e.g., DSL, cable modem, dial-up, wireless, fiber) providing Internet service to specific geo-coded (i.e., by Latitude and Longitude) residential addresses; 2) web-based and telephone research, including address-level service-availability queries of web sites operated by service providers and independent entities. This multi-sourced SNG dataset is used as a validation source for provider service area coverage, Technology of Transmission, and Speed.

American Roamer Wireless Market Intelligence Data: Commercially available dataset used as an independent source to verify information submitted by Providers of wireless broadband service. This dataset is used as a validation source for provider service area coverage.

Strategic Networks Group (SNG) Targeted Online Surveys: Questionnaires (e-mail/web based) have been sent directly to businesses and households, including over-sampling in rural area and those where the above conservative estimate indicates are "unserved" and "underserved" areas. In addition to collecting broadband supply data on type of access, speeds, price, etc. questionnaires gather broadband service demand and usage data from businesses, organizations, and households. Survey responses include geographic coordinates that allow mapping and cross-reference to census blocks or street segments. This dataset is used as a validation source for provider service area coverage, Technology of Transmission, and Speed.

Online Public Survey and Speed Test: A Broadband Mapping Public Survey Site is deployed. Site visitors are requested to provide data on broadband availability, technology, service type (e.g., speed tier) service provider name; monthly prices paid and measured downstream and upstream speeds. In addition to State promotion via press releases to the general public, the State Council on Post Secondary Education (CPE) also promoting participation on this survey to the faculty and student population. This dataset is used as a validation source for provider service area coverage, Technology of Transmission, and Speed.

Prior Broadband Mapping: Statewide coverage areas for Cable, DSL, and Fixed Wireless providers that were aggregated as part of a previous broadband mapping effort for the Commonwealth of Kentucky are used to validate against Provider submitted data. In addition to the service areas, the DSL and Fixed Wireless layers contain general speed information that can be compared against Provider submitted data.

FCC Speed Test: The FCC speed test data includes the IP addresses for each specific speed test conducted. This IP address is queried against a web search engine to determine the Provider assigned to that address and is used as a validation source for provider service coverage and typical speeds.

**Field Data Acquisition:** Broadband technicians visited a sampling of census block locations to gather broadband data to be used for validation. The following criteria were taken into account when developing the census block sampling dataset:

- urban vs. rural census block characteristic
- census block grouping
- land vs. water census block characteristic

The overarching mission of the Federal broadband stimulus program is to expand Broadband service to areas that are currently unserved and underserved. Also, the market intelligence validation sources typically represent some rural, but more urban areas. Thus, our field data collection efforts were targeted more towards the rural areas; split 90% rural, 10% urban.

Additionally, a study by Penn State University (Glasmeier 2002) notes that a large number of census block groups typically fit within any given cable or telephone company service areas. Therefore, our field sample was also based on selection of one census block per block group. The selected census block also had greater than 50% land area, versus water. There are a total of 3,158 census block groups statewide. Using a statistical sample size calculator based upon the number of block groups in the state and +/- 4% margin of error at a 95% confidence level, the sample size is 529 census block locations (Figure 3).

For the 529 census blocks that were visited, 2455 individual wired/wireless data elements were recorded and 3024 pictures were taken at those locations. This field collected dataset is used as a validation source primarily for wireline and wireless technology of transmission and middle mile, and for wireless speed.

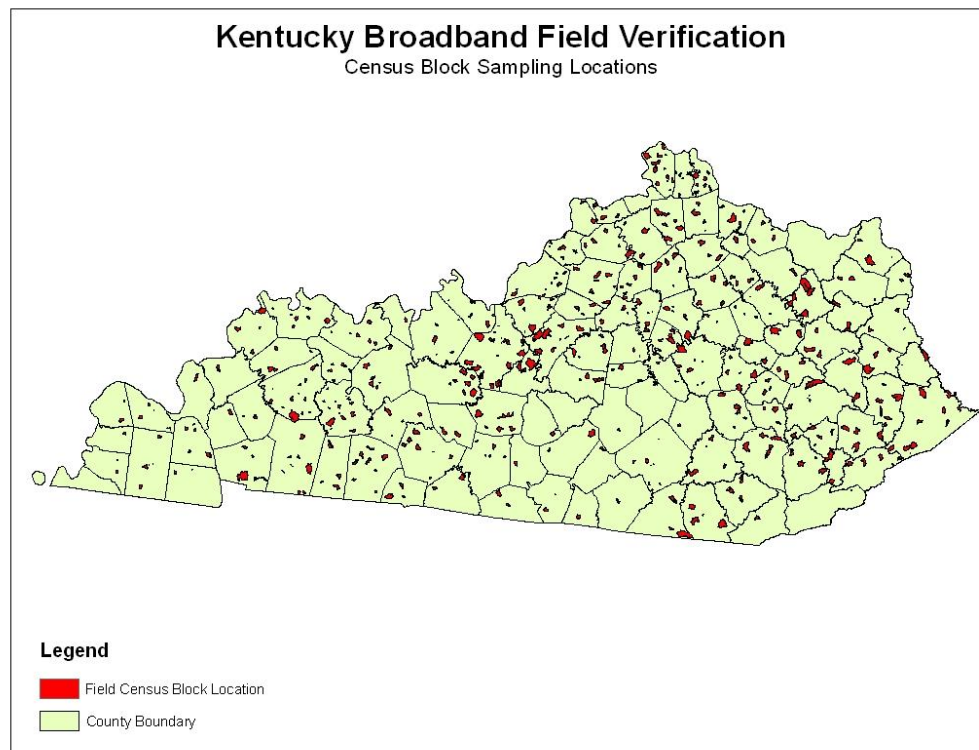


Figure 3 Field Verification Sampling Locations



For each census block in the sample set, broadband technicians collected data using Panasonic Toughbook computers, loaded with MapPoint mapping software, and a customized Microsoft Access data collection form with the ability to automatically import GPS coordinates. The sample census blocks were pre-loaded and directly accessible from MapPoint. Two types of data collection were conducted; infrastructure observation and wireless speed testing; and the results were recorded and linked to the corresponding field location coordinates within the designated sample census block. The information collected by the field broadband technicians includes:

Wireline:

- GPS coordinates
- circuit infrastructure feeding the area (copper, fiber, cable)
- collect site pictures

Wireless:

- GPS coordinates
- internet speed test

This field collected dataset is used as a validation source primarily for wireline and wireless technology of transmission and middle mile, and for wireless speed.

Independent 3<sup>rd</sup> Party Validation: Murray State University coordinated the efforts of resources at the University of Louisville and the Kentucky Community and Technical College System (validation team) to validate the collection methods and collected data associated with the collection of broadband availability data. This validation data developed from this effort was subsequently integrated into the Statistical Evaluation and Assessment System (SEAS) to verify the data submitted by the broadband providers.

The validation team review included:

- a. Validating the list of providers being used by the mapping vendor to make sure all providers are included.
- b. Validating the list of state-provided and Census Tiger Data to identify the location of health facilities, schools, libraries, hospitals, universities, public buildings, etc.
- c. Reviewing provider outreach methodology being used by the mapping vendor.
- d. Reviewing submission options, the Non-Disclosure Agreement and the timeframe for submission.
- e. Identifying Business Intelligence data sources to validate provider information.
- f. Reviewing mapping vendor's website used to collect comment/survey forms from visitors to validate the broadband coverage in their area.
- g. Observing the data collection and data entry process and the ongoing steps in the development of the final products.

Once data was collected, the validation team provided a review that included:

- a. Cross checking of data for accuracy
- b. Statistically representative and significant samples to validate data, especially in rural and potentially underserved.

Limited field census and telephone surveys were also used to validate data in situations where the data cross checks and statistical samples are not able to validate data provided by the mapping vendor. Faculty and students from campuses of the Kentucky Community and Technical College System (KCTCS) conducted the field census work to validate local adoption rates. KCTCS has 16 colleges and over 60 campuses to provide state-wide coverage for field census work.

The work performed, and being performed by the validation team can be summarized in four areas: (1) Audit, (2) Selective Surveys, (3) Reconcile Survey and Provider Data, and (4) Field Test to Resolve Discrepancies.

*Audit* – At the beginning of the project it was decided that the best way to obtain quality data was to make sure that the initial data collection was of the highest quality that it could be. The validation team concentrated its initial efforts in working with the mapping vendor to get the best quality data and also the largest quantity of data that could be obtained. Mapping vendor processes were reviewed and suggested improvements provided. Web sites and documents that were to be used for data collection were evaluated and improvements suggested. Provider lists were reviewed and additional vendors or potential vendors were identified by the validation team. Once data collection began, the validation team also worked with the mapping team to increase the amount of data collected. KCTCS provided web survey sites to students and faculty across the state to increase participation. Once the data was collected the validation team worked to identify data anomalies and locations where additional data collection was required.

*Selective Surveys* – The data audits identified locations where there was insufficient data to make valid conclusions about broadband availability. The validation team used a call center to place selective surveys in the targeted areas within the state. In many cases the insufficient data was the result of the failure of vendors to provide data to the mapping vendor. The selective surveys provide validation of the availability of broadband or the absence of broadband within a specific area. This information allows the mapping vendor to concentrate their efforts to obtain the required data from the appropriate vendor. The call center efforts reached almost 10,000 new households that had not been sampled by other methods. The data indicated that 68.8% had computers, 64.7% has access to the Internet, and 56.7% has broadband access. The new data points were located in rural areas of the state and were focused on areas that had been underrepresented in prior data collection efforts.

*Reconcile Survey and Provider Data* – The mapping vendor survey data (from web surveys), the provider data, and the selective surveys done by the validation team provide an additional reconciliation of the data. While the importance of knowing where broadband is available is critical, it is just as important to know where broadband is not available. The comparison of the various data sources allow for a high confidence in identifying where broadband is available. Additionally, the data reported on the web surveys and the phone surveys identify pockets of citizens of the Commonwealth that don't have access to broadband. The validation team used the data reported by the providers, the data collected by the mapping vendor, and the validation survey data to identify areas of interest for the field data collection efforts. The focus of the field data collection efforts are areas with no reported service, areas where individuals report no availability, and areas where only mobile wireless has been reported as being available for broadband service.

*Field Test to Resolve Discrepancies* – The reported territory covered by wired broadband infrastructure is reliable. However, the reported territory covered by wireless broadband infrastructure (especially mobile wireless) is less reliable. Many factors can impact the availability of the wireless signal. We simply have to think about our cell phone usage and the frequency of dropped calls or no service availability. It is relatively easy for a vendor to say they provide service to an entire geographic area. The validation team developed software to check on the level of mobile wireless availability and to make sure it is at broadband speeds. The validation team drove mobile devices around the state collecting signal strength and doing periodic speed test to validate

the availability of broadband. The initial focus was on areas reported to have no service and areas that only have mobile broadband reported. Test data was collected to validate the data collection process and identify required equipment.

### Provider Data Validation Process

Provider Feedback Loop: Feedback received from the providers is visually inspected and integrated directly into the mapping GIS database.

Service Area Validation Data: The SNG wireline service area data is tabular and contains a separate record for each provider/technology of transmission combination with an associated census block or TIGER road segment, depending on whether the size of the census block area ( $\leq$  or  $> 2$  sq. mi.). This data is exported into an ArcGIS data format. The American Roamer and Prior Mapping service area data is already in an ArcGIS data format. The validation data is then joined to the Provider service area data by census block or TIGER road segment ID. Any database records in the Provider or Validation tables that cannot be joined are output to a separate layer that indicates the areas of discrepancy between the two datasets. The joined tables are then queried to detect any speed discrepancies which are also output to a separate discrepancy layer.

Online Surveys, Field and Independent 3<sup>rd</sup> Party Validation Data: The Public and Targeted Business/Household survey, field and independent 3<sup>rd</sup> party validation data are also collected in tabular database format, and represent a specific lat/long spatial location for each record. This data is exported into ArcGIS data format, joined to the provider data, queried to validate pertinent attribution. Again, records not joined and or with detected attribution discrepancies are output to separate GIS layers.

Topology: The ArcGIS Validate Topology Tool is used to flag any topology issues in the broadband data. Flagged issues are reviewed to identify false positives and update true errors as required.

SBI Check Submission: The NTIA-provided SBI Check Submission tool is utilized to validate that the deliverable broadband data is consistent with the business logic rules set forth by the NTIA and a passing receipt is provided with the data submittal to NTIA.

Stakeholder Feedback: The state broadband mapping website includes a feedback function. Comments received from stakeholders are reviewed and used to validate provider data submissions.

### Validation and Confidence Level Reporting

To facilitate validation and confidence level reporting, Baker deployed a validation application called Statistical Evaluation and Assessment System (SEAS) which automatically compares the multiple independent validation datasets against the broadband service provider supplied information. The SEAS uses statistical methodologies to report the confidence level in the spatial and attribute accuracy of the information. Appendix B shows the validation workflow.

The SEAS comparison is a three-part validation process:

1. Comparison of the collected validation source against the aggregated broadband provider data.
2. Match percentage calculation for each provider reported in the DataPackage.xls, "Provider Table" tab, "Comments" column.
3. Confidence score calculation displayed on the state broadband website.





Figure 4 Statistical Evaluation and Assessment System (SEAS)

After completing all validation data source collections, SEAS is used to automatically compare the multiple validation datasets against the aggregated broadband data which came from the providers. Through the SEAS accumulation table, it produces a match percentage per broadband service record based upon the number of matches that record has against each validation source. The matched percentage for each record is the result of the total count of the matched validations for the record divided by the total validation source being compared against the record. A validation confidence rating/score is then assigned on a scale of 1 to 5 based upon the percentage of validation source matches as per the following score results:

- 1 Star = 0% - 19% Match
- 2 Stars = 20% - 39% Match
- 3 Stars = 40% - 59% Match
- 4 Stars = 60% - 79% Match
- 5 Stars = 80% - 100% Match
- “No Analytics” = No validation source available for that provider

The Commonwealth's public broadband mapping website ([www.bakerbb.com/kybroadbandmapping/](http://www.bakerbb.com/kybroadbandmapping/)) is updated with the confidence level results at the record level based upon the queried geographic location and the following is an example of this representation.

Provider Name	Transmission Technology	Max Download Speed	Max Upload Speed	Confidence Score
AT&T Mobility	Mobile Wireless	Greater than or e...	Greater than or e...	
Verizon	Asymmetric xDSL	Greater than or e...	Greater than or e...	NO ANALYTICS
Comcast	Cable Modem – Other	Greater than or e...	Greater than or e...	

The matched percentage for the records for each provider are summarized and then divided by the total count of the records to create the final matched percentage for the specific provider. These percentages are included in DataPackage.xls on the Provider Table tab in the Comments column.

### Low Confidence Provider Feedback

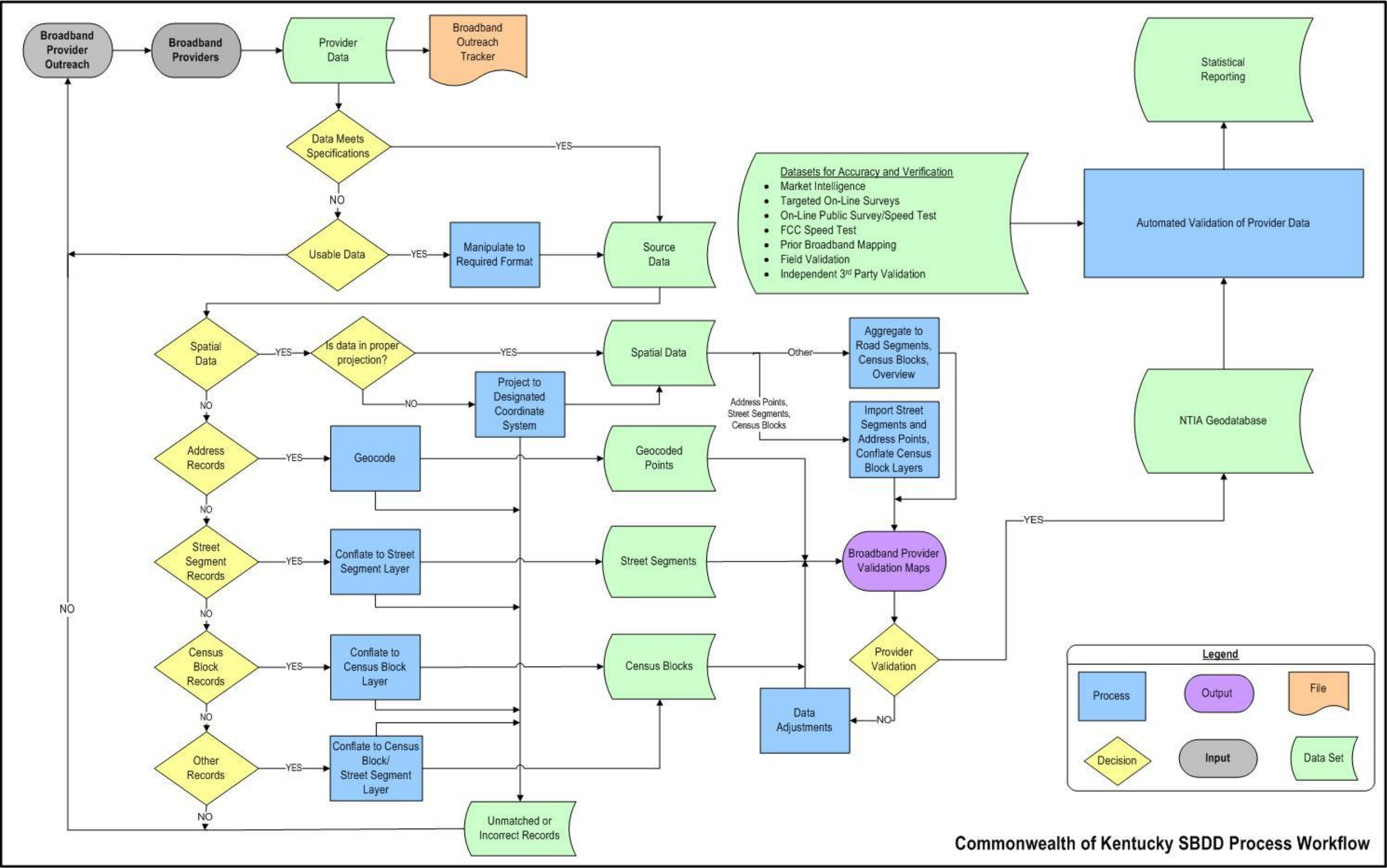
Provider data which is assigned a low confidence (1 or 2 stars) through the SEAS process is communicated back to the provider through a feedback loop. Generally, the low confidence feedback and reconciliation is a continuous refinement process and will occur between update cycles. The goal is to provide this feedback through the Provider Update Portal via a web connection that will be available and rolled out to provider in early 2012.

### Changes and Corrections Documentation

With each NTIA semiannual data submittal, changes and corrections documentation is provided. Significant changes in a provider's status or data, corrections to previously supplied data, providers supplying data for the first time, etc. are specified by Provider name in the Changes and Corrections document.

### Future Validation

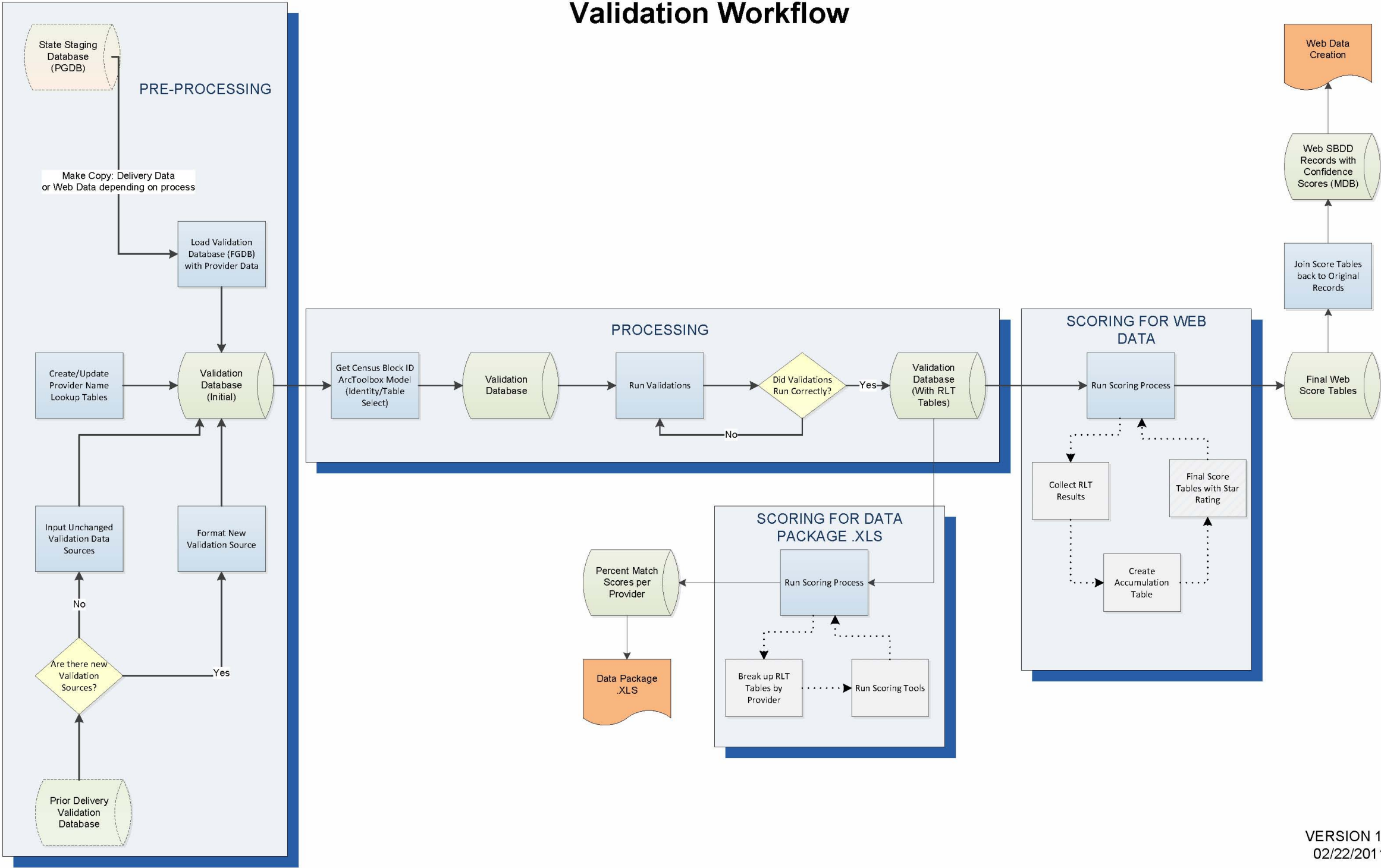
Audit of Wireless Broadband Availability Reporting: Wireless coverage will be evaluated using a contour calculation tool, with key inputs being transmitter location and, where available, data on spectrum power levels and other relevant transmission factors provided by carriers and/or supplemented by data available from public web sites and other sources. Data will then be input to a contour calculation tool to provide estimates of fixed wireless broadband coverage areas. This dataset is used as a source to determine gaps in provider wireless service area coverage. The Prior Mapping data is also used as a validation source for gap analysis.



Commonwealth of Kentucky SBDD Process Workflow

October 1, 2010

State Broadband Data  
Validation Workflow





2000 to 2010 Census Data Translation Workflow

