NC Broadband Information Submitted October 10, 2010

To: NTIA Mapping and Planning

From: Stephanie Jane Edwards
Data Administrator/GIS Specialist

Through: Jane Smith Patterson
Executive Director
The e-NC Authority

November 5, 2010

General Information Regarding This Data

One hundred and fourteen providers (by DBA) of broadband service were identified to be currently operating in North Carolina (up from 95 in the spring), and of these a total of 56 providers submitted some form of broadband data, either complete or partial, through the process of communication and technical support from e-NC (See DataPackage.xls). All providers were requested by email letter and hard copy to provide data to our collection process, and efforts were made to reach the providers by phone as well.

A Few Technical Notes Regarding This Data

When a geocoding process was necessary to use a geocode process or spatial join to features from other features or records associated with the reported data, the fields from reported data were always selected when loading the resulting data into the geodatabase feature class. This can create situations where, for example a particular address is reported to have service availability and that address is the exact one reported even though the point feature may have originally been that of an address three houses down the street. This prioritizes trueness to the actual broadband data reported rather than changing values so that they match the reference geometry used 100% of the time.

Data Not Found in this Delivery

Epproach Communications, Inc. Middle Mile: Epproach Communications did submit to e-NC their information on middle mile, which for them consists of a single connection point. However, this could not be included because they reported (in Excel spreadsheet) the same value for both latitude and longitude, and have not yet responded to inquiry about the missing coordinate.

Data for provider Zito Media: Zito Media is a very small provider serving in two census tracts of Graham and Swain Counties, and was only able to initially send PDF maps showing these two census tracts, though the actual area served is thought to be much smaller within these, and therefore e-NC is working with Zito Media to produce more specific information rather than over-reporting the served area.

Data for provider Opterex: Opterex reported data in the spring for a single tower of licensed wireless broadband service to some departments of municipal government only in Leland, NC. This data is not included in this Fall 2010 data package because Opterex has dismantled this tower and now utilizes AT&T Mobility's 3G network to create an encrypted mesh network. Their modifications to the signal and mobile receivers in public safety and public works vehicles provide a custom solution to these local government entities. End user locations include only municipal-owned vehicles, creating an entirely mobile context. It was determined by the e-NC Authority that no additional data relevant to this mapping project could be collected from Opterex for North Carolina, and the data that is relevant has been collected from AT&T Mobility.

Satellite data: Per instruction by the NTIA planning and mapping team via webinar, satellite data was excluded from this data delivery. One satellite broadband provider in NC reported information to e-NC, and it was a solid coverage representation over the entire state, which e-NC expected but considers a generalized overstatement of actual availability (particularly for areas in valleys and on the northern slopes of mountain features).

Integration of Provider Data into NTIA Statewide Geodatabase

For ease of data integration, a front-end Excel format template was offered to all providers, containing notes defining required fields, explanations of which data is required in which formats by which types of providers, and hyperlinks connecting fields to additional tables listing the corresponding NTIA-specified values and codes (for speed tiers, technology types, connection point facility types and capacities, county codes, end user types). A brief description of how census block FIPS codes work was also taken from an internet source and distributed as needed to providers who had questions about how to report this information.

BB Service by Census Block: As requested by the NTIA mapping and planning team, all census block data is included with 2000 census block geometry. All relevant data was reported as such by providers in North Carolina, except for Time Warner Cable and Charter Communications. These datasets were translated into 2000 Census block geometry by two different processes; it was found while working with the Charter data that a process involving specific selection by location followed by a spatial join produced results that eliminated the significant overestimation resulting from a direct spatial join of the dataset with 2000 census block geometry using "intersect", and the dramatic underestimation using "contain". This new method will be tested further and the resulting improvement quantified in terms of estimating margin of error and thus justifying a best practice. Technical assistance was often needed by providers to correctly report served areas by either the 15-digit FIPS codes or in some way by which e-NC staff could derive the appropriate FIPS codes.

BB Service Road Segment - The reporting and mapping of data by street segment presented significant challenges to accurate interpretation of where broadband availability is and is not. This is mainly attributed to the difficulty of standardization among the many data structures by which providers report street segments.

BB Service Address – A few address-level datasets were submitted to e-NC with the latitude/longitude coordinates already included, but most needed to be geocoded. This was done using the NC Master Address file as the primary reference file, significantly increasing the accuracy of matching records. Secondary sources for address records that did not find a match this way included street segment interpolation, ESRI data utilizing the 4-digit ZIP extension, and manual placement/digitizing based on a combination of reference data and online browser maps. Upon completion of geocoding for each provider submitting address data, the address point features were overlain with a 2000 census block layer to add the census block FIPS code attribute, then all relevant features were loaded into the geodatabase feature class. Some duplicate addresses were discovered besides those addresses served by multiple providers, but not easily removed with the setup prior to fall submission. For example, Mediacom reported over 108,000 flat-file address records, and this data contained 3,865 duplicate addresses which were not removed prior to geocoding; once it was noted that address data may contain such duplicates, an extra quality control step was added to remove duplicates prior to geocoding from within a given dataset. Due to aggressive timelines/time constraints on available staff, some address-level data was geocoded and loaded before a maximum match percentage could be attained, which means that fewer addresses at which broadband is available are indicated by the address feature class contents. The geocoded shapefiles for each provider are kept with geocode match score and match reference type for every matched address, so the thoroughness of this data type could be tracked and/or improved with more time.

BB Service Wireless - Approximately seven small, fixed wireless providers have been able to share technical information about their transmitting towers, antennae, and frequencies, so that e-NC can produce for them a service coverage shapefile using the contracted services of the University of NC at Greensboro Center for Geographical Information Science.

BB Service Overview – Records for overview containing subscriber-weighted nominal speeds of a given provider were generally joined to a template layer of county features, using the option to keep matching records only. Then these matching features and their new attributes were exported as a new shapefile before being loaded into the collective overview feature class.

BB Service - Critical Anchor Institutions – Only anchor Institutions that could be geolocated were included, and there were only 17 CAI's that were identified and could not be geocoded to a point feature. These were collected by contacting administrative offices of some CAI category types and receiving databases of information, as well as collecting from individual CAI locations for other types using survey emails and follow up phone calls as necessary. There are 4,670 CAI's identified, located, and included in the geodatabase to date.

Verification Implemented Prior to Fall Data Submission

Data verification methods that were implemented by e-NC in time for submission at the federal level were generally along the lines of quality control. These are outlined below. Time constraints on existing

staff did not allow for the execution of some less basic verification approaches which are described further below and are in the planning/setup stages.

Standardizing – The files from datasets received from each provider, except for those few submitted in shapefile format, were manually transferred to a back end Excel-format template with field headers, to create a single-file, standardized field structure for each provider's data that could be used for quick reference and map feature creation. This step also helped staff to ensure that all required components were either present or requested in follow up to the provider, and that the components were reported in the correct format.

Lat/long coordinates – Some information was submitted to e-NC with lat/long coordinates included for the location of point features. This location information was checked during the mapping process, and values were corrected if the provider had made mistakes such as reversing the latitude with the longitude, or forgetting to include the negative sign for the longitude value. In addition, e-NC followed up with providers on point features that showed up in the map outside the state and/or outside the provider's reasonably expected service area. Point features that mapped outside the state after follow up with providers, including those that mapped to zero degrees latitude and longitude due to an unknown location, were deleted from the geodatabase for submission at the federal level. For fixed wireless data generated by propagation model from antenna specs, the latitude/longitude coordinates of the antenna locations reported by the provider to e-NC were verified by e-NC's university GIS research contractor using high-resolution orthoimagery.

Multiple FRN's – In several instances, providers reported multiple FRN's that increased in numerical increments of one for each record of data, and this was found to be a simple error when the providers were trying to paste their organization information down the rows applying to a list of broadband data records. This was checked for and corrected after confirming that the lowest/first reported FRN was the correct one.

Correct technology type codes – Knowledge from our technical staff and online research was sometimes used to supplement data that e-NC had relevant to a provider that was unresponsive or otherwise did not supply this specific piece of the information. For example, a provider may have gaps in their transmission technology field and these were filled in when technical staff could confirm that the provider operates with only a single technology type. Or the staff may know which technology type is used by a provider who simply left this field blank on all records.

Subscriber-weighted nominal speeds – Weighted nominal speed values were checked, and staff followed up with the provider if all values were the same for multiple counties, as this could result from either a single speed tier for a given transmission technology across counties, or in some cases providers were not following the formula provided and had manually entered the same value regardless of differences in subscriber numbers. When these cases were discovered, technical assistance was offered and a new subscriber-weighted nominal speed dataset created to reflect variation between counties.

Wireless model fieldwork – For fixed wireless provider data that was generated as coverage area output from models based on technology and environmental factors, the data was verified by "ground-

truthing" with measurements of signal strengths at sample locations within a provider's service area, observation of the influential ground conditions in each location, and comparison to the expected signal strengths at the same locations in the model. Some calibration of the model was then performed so that the resulting polygons could more accurately reflect what would be found in real life.

Check geometry – After compiling all datasets into the geodatabase feature classes, the check geometry process in Arc Toolbox's Data Management section was used on each feature class to identify and repair any geometry errors in the features.

Planned Verification Upon Completion of Setup

Many hours and resources have already been invested by e-NC into conducting and evaluating the effects of multiple data collection methods (besides direct collection from providers) for use as alternatives, supplemental methods, and/or verification of provider-sourced data. These alternative methods have been described in proposals and quarterly reports and primarily consist of: citizen and consumer sample surveys, web crawling/commercial databank-sourced, and the empirical propagation model for fixed wireless data. Pending completion of an SDE-in-PostGreSQL database on a secure Linux server will facilitate cross-method and cross-data pull comparison.

Establishment of the SDE database setup will also allow for faster robust queries and analysis combining SQL with the mapping environment, as well as the addition of some more business rules to automate the flag of data that contradicts expected values (such as known/expected service area extents by provider). We also seek to pursue the streamlining of update processes and results via automated removal of known duplicates between newly and previously submitted datasets. This should significantly reduce redundant geocoding and mapping hours required. Techniques and lessons learned from the upcoming master address file improvement project will support this aspect of the work.

The e-NC Authority also hopes to add in crowd-sourced speed test data and reports of no access, currently being collected from e-NC's website into a database, for an additional verification source. The speed test used by e-NC incorporates a server site in North Carolina's Research Triangle Park, thus providing more accurate speed data for end users in the state. The FCC speed test results provided on the federal "State Broadband Data Management" website may also be utilized for comparison since it is available. The online map viewer for NC will also feature feedback comments about the data from map users such as individual citizens, representatives, and broadband providers. Providers will also have their own login-access view of their specific service area on which their comment can be made as to the accuracy of their reported information and/or e-NC's interpretation of it.

In Consideration/Exploration Stage – The possibility was recently initiated to create a feedback loop/added verification from NC county-level government groups that are working on broadband access improvement in their counties and who may receive data from other sources or conflicting data from the same provider. Wireless model ground-truthing could also be used for verifying data submitted as shapefiles from larger mobile wireless providers.