The U.S. National Broadband Map: Data Limitations and Implications

Tony H. Grubesic, Geographic Information Systems and Spatial Analysis Laboratory, College of Information Science and Technology, Drexel University, Philadelphia, PA 19081.

Email: grubesic@drexel.edu

Abstract

The 2011 release of the National Broadband Map (NBM) has generated significant interest from the telecommunications policy community. The NBM is a multiagency effort, including the National Telecommunications Administration (NTIA), the Federal Communications Commission (FCC), fifty U.S. states and five U.S. territories, to collect and disseminate information on broadband provision and quality of service for the United States. From a geographic perspective, the NBM represents a marked departure from previous broadband data efforts. Specifically, instead of disseminating FCC Form 477 data on providers at the ZIP code or Census tract level, the NBM reports provision information at the Census block level – the smallest geographic unit in which the Census bureau tabulates survey information. While this increased level of geographic data resolution is a welcome change, there are several notable limitations to these data that are important to consider when conducting spatial-econometric analysis for public policy evaluation. With this in mind, the purpose of this paper is two-fold. First, this paper explores the salient characteristics of Census block geographic base files, highlighting their strengths and weaknesses as summary units of spatial analysis. This also includes a brief discussion on how the NBM treats geographically large blocks and their use of road segments for aggregating provider data. Second, this paper examines the impacts of “empty blocks” (i.e. blocks that have no household information associated with them) for the spatial analysis of broadband. Finally, this paper provides a short overview of how these data limitations can impact public policy evaluation and provides a roadmap for improving the National Broadband Map.

Keywords: Broadband, National Broadband Map, Spatial Analysis, Econometric Analysis, Public Policy, Geographic Information Systems.

UNDER REVIEW: PLEASE DO NOT REDISTRIBUTE WITHOUT PERMISSION
1. Introduction

February 17th, 2011 marked the release of the National Broadband Map (NBM) by the Department of Commerce’s National Telecommunications and Information Administration. The NBM is the first searchable nationwide map of broadband Internet availability made publically available for the United States. According to Rebecca Blank (2011), the Acting Commerce Deputy Secretary, “[t]he National Broadband Map, will inform efforts to enhance broadband Internet access and adoption – spurring greater innovation, economic opportunities and advancements in health care, education and public safety”.

While there is no doubt that the high resolution data provided by the NBM represents a marked departure from the existing data used to evaluate local broadband options, namely the Form 477 database from the Federal Communications Commission (FCC) (Strover, 2003; Grubesic and Murray, 2004; Grubesic, 2006; Grubesic, 2008a; Kolko, 2010), there are a number of issues associated with data integrity, spatial uncertainty and accuracy within the NBM that need to be addressed. For example, a recent report suggests that the NBM does not include all broadband carriers or provide an accurate portrayal of coverage. Specifically, IDInsight (2011) notes a wide variation in participation rates for broadband providers in each state. For instance, Virginia had the lowest participation rate, where only 27% of providers (36/132) engaged in the NTIA data collection process. Clearly, this makes an assessment of broadband availability for Virginia nearly impossible. In fact, fourteen states had provider participation rates under 60%.1 In contrast, there were eight states that had 100% participation.2 In a more pointed critique of the NBM, Lennett and Meinrath (2011) and Meinrath (2011) suggest that the reliance of the NBM on self-reported provider data is highly problematic. First, providers “often paint their coverage areas with a broad brush” (Lennent and Meinrath, 2011, 2). Second, the NBM merges

1 ME (58%), KY (58%), WV (57%), FL (54%), NC (53%), RI (51%), DC (49%), AL (48%), LA (47%), NM (47%), MD (41%), CO (39%), MO (35%), VA (27%).
2 IN, IL, MS, CA, DE, GA, NH and HI.
business and residential services. As a result, while some areas may appear to have a plethora of service options, the majority of providers are targeting businesses, not private residences.

Beyond these fundamental problems associated with data collection and interpretation, a myriad of traps and pitfalls exist for analysts interested in using the raw NBM data for empirical evaluation. In addition to the basic computational challenges associated with working with over 12.5 million data points, developing the appropriate aggregation routines to re-scale the data for econometric analysis and policy evaluation is not straightforward. For instance, raw data were collected by the NTIA (and subsequently provided to the general public) using three different files: 1) if the block is less than two square miles, provider information is tabulated by block number. 2) When blocks are two square miles or greater, provider information is tabulated by street segment. 3) Wireless data were provided to the NTIA as geographic coverages (i.e. digital maps of wireless service coverage) which were then overlayed on top of census blocks to calculate the percentage of each block covered. Developing a more thorough understanding of how these tabulation approaches might impact broadband analysis is critical to the development of meaningful economic development strategies and associated broadband policy.

With these data challenges in mind, the purpose of this paper is to empirically illustrate the strengths and weaknesses of the data associated with the National Broadband Map. Admittedly, this is not intended to be a comprehensive review of all aspects of the NBM data because there are far too many components of these data to cover in a single paper. However, within this general framework, two specific facets of these data will be addressed, emphasizing their spatial nature. First, this paper provides a brief evaluation of the critical differences and associated analytical impacts of broadband data tabulated for large blocks (> 2 square miles) and small blocks (less than two square miles) for the NBM. Second, this paper explores several geographic nuances of Census blocks, examining their spatial structure and the implications of empty blocks (i.e. administrative units that have no household
information associated with them) on local spatial statistical analysis. Policy implications and strategies for dealing with data uncertainty are also reviewed.

2. Broadband in the United States

In March 2010, the Federal Communications Commission (FCC) released the national broadband plan (the Plan) for the United States (FCC, 2010a). The Plan outlines a multifaceted and highly complex strategic agenda for developing and enhancing broadband infrastructure. In part, the goal of improving U.S. broadband infrastructure is related to the simultaneous desire for improving health care, education, energy use, economic opportunity, government performance, civic engagement and public safety through the more fluid and dynamic exchange of information and knowledge via broadband networks. Consider the following statistics. The FCC (2010a) determined that nearly 100 million Americans (32%) do not subscribe to any form of broadband service. Similarly, OECD (2010) statistics suggest that the United States has 27.1 broadband subscriptions per 100 inhabitants (OECD average = 24.2) and ranks 14th globally. The OECD statistic is particularly interesting, considering that the top-ranked countries such as South Korea and Denmark exhibit 37.8 and 37.3 broadband subscriptions per 100 inhabitants, respectively. With the recent allocation of $7.2 billion in federal stimulus funds to improve broadband in the United States, it is clear that the federal government and its policy makers believe these broadband gaps require mitigation (ARRA, 2009).

These statistics, however, only hint to the depth and complexity of broadband provision and access issues in the United States. For example, In July 2010, the FCC (2010b) reported that between 14 and 24 million Americans are incapable of getting a broadband connection to the Internet. Clearly, this is a far cry from the 100 million Americans that do not subscribe to broadband. Phrased somewhat differently, these statistics suggest that between 4.5 and 7.7 percent of the U.S. population does not have access to broadband. To put this in perspective, as of June 2011, over 7 percent of U.S. counties
are not capable of providing enhanced 9-1-1 service to their residents, where emergency calls are routed to public safety access points with information regarding the caller’s phone number and location (NENA, 2011). While one cannot compare these statistics directly, it is clear that the provision of telecommunications services, whether broadband or enhanced 9-1-1, is challenging in the United States, yet where broadband is concerned, the vast majority of the country (> 92%) has access.

Unfortunately, our understanding of broadband provision, access, use and its related economic and social impacts remain obscured for three reasons. First, much like the digital divide, broadband is a moving target. Because wireline and wireless broadband technology platforms are constantly being modified and improved, the availability, price, performance and quality of service associated with these technologies is continually shifting. Second, the relatively competitive local, regional and national markets for broadband guarantee a highly dynamic environment for many providers and subscribers. Finally, from a historical perspective, the U.S. has suffered from a dearth or quality broadband data (Greenstein, 2007). For over a decade, the only comprehensive database for exploring broadband service provision in the United States was supplied by the FCC as Form 477 data. The limitations of these data are widely noted elsewhere (Flamm and Chaudhuri, 2005; Grubesic, 2006, Prieger and Hu, 2008; Grubesic, 2008b), with serious problems in both their geographic coverage and the service details collected from each provider.

One positive result from the growing emphasis on developing broadband capacity and accessibility in the United States is the National Broadband Map. The NBM represents a massive synthesis of survey and research effort performed by the NTIA and the FCC in partnership with fifty states, five territories and the District of Columbia. Its primary goal is to provide a biannual, high-resolution snapshot of broadband availability in the United States. While broadband provision data are collected and verified by state-level agencies, they are compiled and made available via the NBM through efforts by the NTIA. As noted in the introduction, full participation in this effort was not
achieved.\textsuperscript{3} Specifically, Of the 4,600 broadband providers identified nationally, data for 3,400 (~74\%) were collected for the NBM (NTIA, 2011a). Further, because of the number of state-agencies involved, the diversity associated with broadband carriers (size and platform), and the geographic extent of their service areas, data collection and tabulation routines for each state were not always identical, although the final integration of the broadband information was done through a specific data model.\textsuperscript{4} As a result, each state provided a unique data collection methodology that is available to interested parties.\textsuperscript{5} Thus, although Steven Rosenberg, the chief data officer with the FCC’s Wireline Competition Bureau touts the NBM as the “largest and most detailed map of broadband ever created” (Meinrath, 2011), it is far from perfect.

3. Wireline NBM Data

While the core data elements for the NBM are associated with broadband provision (e.g. providers and download/upload speed), this information is tied to Census-defined tabulation units known as blocks. Census blocks are the smallest geographic area for which the U.S. Census Bureau collects and reports decennial survey information. Blocks are typically bounded by local features such as streams, streets, highways and railroads, although cultural features such as schools, prisons and other major features also impact boundaries. For Census 2000, data were tabulated for 8,262,363 blocks for the fifty states, Puerto Rico and the District of Columbia.

From a geographic perspective, there are several notable aspects to census blocks worth mentioning. First, a census block is always unique to, and can never cross the boundaries of census

\textsuperscript{3} Neville (2011) notes that the top 10-15 providers by broadband technology type are included for each state in the NBM. She suggests that this comprises close to 95 percent of the broadband market, even though there are some providers that have not been included.

\textsuperscript{4} http://www.broadbandmap.gov/about/technical-overview/data-model

\textsuperscript{5} http://www.broadbandmap.gov/about/state-broadband-programs
tracts or block number areas (Census, 2000). This facilitates a nested, geographic hierarchy that allows for reasonable aggregation schemes to be developed between small area data and larger administrative units (e.g. block groups and tracts) where more detailed socio-economic information is available. Second, while most major urban locales in the United States maintain a high density grid of relatively small census blocks, more recent suburban and exurban development includes curvilinear street patterns, cul-de-sacs and larger expanses of formal greenways. As a result, the spatial profile of blocks in these areas tends to be larger, often reflecting both branching and grid-like patterns - depending on local settlement geographies. Figure 1 highlights these variations in block morphology for portions of Columbus, Ohio and its suburb, Dublin. While much of Columbus adheres to the grid structure, Dublin is more irregular.

As noted earlier, the NTIA collected wireline broadband data in two different ways for blocks. For blocks that were less than two square miles in size, providers simply reported block numbers where broadband service was available. For blocks larger than two square miles, providers collected and submitted address data or road segment data where broadband service was available. Where the large blocks are concerned, the NTIA tabulated instances of service and assigned them to each block. For example, Table 1 displays a subset of NTIA reported provider records for a “large” block, #390490080001000, in Ohio. In this block, there are only two providers present (Time Warner and AT&T), yet there are 55 provider entries in the NTIA database. Again, this likely corresponds to 55 address points and/or road segments where AT&T, Time Warner, or both are providing service.

From a database perspective, there is little difference between small block tabulations and large block tabulations for the NBM. Aside from minor variations in the fields included, the core data

---

6 Block numbering areas are small statistical subdivisions of a county for grouping and numbering blocks in non-metropolitan counties where local census statistical area committees have not established census tracts (Census, 2000).
7 10 of the 55 records from the NTIA database are illustrated in Table 1. All of the records reported in the NTIA database for block 390490080001000 were associated with Time Warner or AT&T.
components (provider name, upload and download speeds and platform type) are reported identically. One outstanding problem, which will be addressed in the next section, is whether or not the reported presence of a broadband provider in a block is truly indicative of service availability. Again, as noted by Lennent and Meinrath, 2011), providers can be overly generous in reporting their broadband coverage, resulting in a drastic overestimation of broadband availability and provider choice for many regions.

A few final notes about the NBM data collected at the block level relates to the survey information reported by the Census Bureau for these administrative units. First, NBM data are linked to Census blocks from the year 2000. This is rather unfortunate, but the data collection efforts for the NBM began prior to the release of the 2010 Census block files. Second, because blocks are quite small in geographic extent, privacy concerns and the fear of revealing personally identifiable information limits what the Census Bureau reports for each block. For example, while data relating to population and household counts, age, race and housing characteristics (rent v. own) are reported at the block level, information regarding income, education levels and occupation are only available at the block group level and higher. The lack of a comprehensive suite of demographic, socio-economic and housing data at the block level necessarily forces analysts to aggregate NBM block tabulations to administrative units where such data are available. Also, because the Census Bureau provides a complete geographic tiling of blocks for the United States, many of them correspond to uninhabited areas. For example, some blocks represent interstate highway exchanges, water bodies or railroad switching yards. That is, instead of these features simply functioning as boundaries to inhabited blocks, these features are blocks unto themselves. From an analytical perspective, providers should not report these “empty blocks” as locations where wireline broadband is available because there is rarely a need, for example, to provide service to a river or lake. Of course, wireline broadband service within a water body is not completely out of the question, just highly unlikely. Regardless, the use of blocks and their associated demographic and socio-economic data availability makes the development of a meaningful and repeatable
aggregation strategy for NBM block tabulations critical for conducting robust and reliable analyses of broadband provision in the United States.

4. Broadband Data Uncertainty and Associated Analytical Considerations

This section provides two simple empirical examples of how data uncertainty can potentially impact spatial-econometric and policy analysis for broadband provision in the United States when using the NBM data. It is important to note that while these problems are not crippling, care needs to be exercised when both preparing and analyzing the raw NBM data.

4.1 Broadband Provision Uncertainty

As noted in Section 3, broadband providers are often relatively generous in estimating wireline service coverages in their market areas. From a technological perspective, operational issues associated with the provision of cable and xDSL are quite different, and far too complex to detail in this paper. However, recent work suggests that the quality of service associated with xDSL provision is highly contingent on geographic location (Grubesic et al., 2011), as is the ability to cover local business and residential demand (Grubesic, 2008b). Specifically, in previous work dealing with FCC Form 477 data at the ZIP code level, Grubesic (2008b) noted the problems associated with assuming that xDSL enabled central offices can provide broadband to an entire telephone exchange area (i.e. wirecenter). They rarely do. This is a mistake that has been made in previous studies (Rappaport et al., 2003), but one that should be avoided at any expense. Further, the core finding of the Grubesic (2008b) piece is that xDSL coverage is not ubiquitous throughout a wirecenter service area for two reasons. First, the geographic range of xDSL service is often limited by line conditions, with high quality service only available at distances less than 18,000 ft. Second, even if a residence or business is within 18,000 ft. of a central office, unless the CO is located within their wirecenter service area, broadband service from the non-native CO cannot be extended across wirecenter boundaries. These results had major implications
when estimating broadband availability at the ZIP code level. Thus, one outstanding question with the new NBM data is if these geographic service constraints are still relevant for the analysis of broadband now that the small area data (blocks) are available for analysis.

Figure 2a displays the Columbus, Ohio metropolitan area, which consists of eight counties, 110 wire center service areas and 130 central offices (CO). Specifically, the CO database consists of all facilities in (and nearby) the Columbus metropolitan area that are listed by the Local Exchange Routing Guide from Telcordia Technologies. This includes information regarding the physical location of COs, the geographic extent of their coverage areas (i.e. wirecenter service areas) and general information on CO capabilities. Figure 2b highlights all of the “large” blocks (> 2 sq. miles) in the Columbus region that have at least one broadband provider in the NBM database. Also displayed are the 18,000 ft. service zones around each CO to estimate the “best case” scenario for xDSL coverage in each wirecenter. Notice that the service zones are clipped, so that there is no overestimation of xDSL service range into adjacent wirecenters. Their irregular shape is due to the use of network distance instead Euclidean distance for the estimations of service range – an important consideration when modeling xDSL coverage (Prieger and Hu, 2008; Grubesic, 2008b).

Recall the example of block #390490080001000 used previously to illustrate NBM provider tabulations. Again, in this block, there are only two providers present (Time Warner and AT&T), yet there are 55 provider entries in the NTIA database because each instance corresponds to one of 55 address points or road segments where AT&T, Time Warner, or both are providing service. Figure 3 highlights the location of this block and its spatial relationship to nearby COs and their associated “best case” xDSL service coverages for the region. Notice that block #390490080001000 is located well outside of the 18,000 service range of its central office, although its neighboring large blocks are partially covered. Where xDSL provision is concerned, this suggests that providing service to block

---

8 Small blocks with at least one provider have been omitted from the map to enhance visual clarity.
#39049008000100 is difficult. Specifically, unless there is some type of remote DSL concentrator that is aggregating xDSL customer feeds and routing these data back to the central office (Grubesic and Horner, 2006), the distance constraints associated with routine xDSL coverage (e.g. 18,000 ft. from CO to premise) makes coverage highly unlikely. That said, the NBM reports 18 address points or street segments within this block receive asymmetric digital subscriber line (ADSL) service from AT&T of Ohio.

From a more aggregate perspective, of the 202 large blocks in the Columbus metropolitan area, 72 (~35%) are located completely outside the 18,000 ft. service range of a central office. In addition, only 59 of the 133 remaining large blocks that intersect CO service areas have greater than 50% of their geographic area within service range. Finally, the NBM reports 677 instances of xDSL coverage (e.g. street addresses or road segments) in the 72 large blocks that do not intersect the 18,000 ft. service range of COs for the Columbus metropolitan area. Again, this suggests the broadband availability for the Columbus, Ohio metropolitan area, as reported by the NBM, is extremely optimistic. From the perspective of population covered, this empirical analysis suggests that 8,829 people reported to have xDSL service in large blocks likely do not — representing a 46% overestimate. It is also highly likely that data for other regions in the United States also suffers from the problems identified in this section.

That said, it is important to acknowledge that remote DSL concentrators are a feature of the broadband landscape in the United States and the statistics detailed in this section may represent a relatively conservative estimation of local DSL coverage for the Columbus metropolitan area. Grubesic and Horner (2006) provided a detailed discussion of the logistical challenges associated with extending DSL coverage to peripheral suburban and exurban areas. Most notably, the costs associated with acquiring rights-of-way to extend fiber connections to remote concentrators can be quite high. As a

---

9 Data from the NBM indicate that this region is served by AT&T ADSL.
10 AT&T’s U-verse is available in portions of Central Ohio. U-verse is a hybrid coaxial-fiber system that offers VDSL (very high bit-rate digital subscriber line) to customer premises.
11 According to 2009 population estimates (ESRI, 2009), 8,829 people live in large census blocks that do not intersect xDSL service ranges for the Columbus, Ohio metropolitan area. A total of 19,173 people live in all large blocks for utilized for this analysis.
result, providers carefully evaluate the tradeoffs associated with these costs and the potential returns on investment for serving lower density suburban and exurban regions. Further, because of the capital costs associated with the installation of remote DSL concentrators, it is highly unlikely that smaller, rural telecommunications markets feature this type of system. The spatial distribution of subscribers is too diffuse and the distances required to connect the remote concentrators with central offices are too large. Thus, while there is no doubt that remote DSL systems likely mitigate coverage gaps in U.S. suburbs, issues of DSL coverage for the national broadband map merit attention.

4.2 Empty Blocks

A second facet of the NBM that deserves attention is the prevalence of empty blocks and reported broadband availability. Recall that empty blocks are locations where there is no population, households or businesses present. The NTIA and FCC are certainly aware of the empty block problem in the context of measuring broadband availability. In fact, for NBM wireline data, all 2000 year census blocks that have zero population are queried against the state broadband data and development (SBDD) information supplied to the NTIA. Specifically, this validation process is used to show that a provider may be overstating their service area in certain instances. For example, if a provider reports that a block is served, but there is no population in the block, the record may be flagged as an error (NTIA, 2011b). The NTIA (2011b, 27) is quick to point out, however, that “this type of data information can be misconstrued because a provider may claim to serve a Census block with 0 population because they can provide service within 7 to 10 days.” This is a good point and one worth discussion. Decennial Census data tabulate nighttime, residential population for blocks, not the daytime (i.e. ambient) population associated with places of business or related commercial districts. As a result, determining truly empty blocks (e.g. those devoid of businesses) is difficult. Therefore, it may be possible to provide broadband service to 0 population blocks or those that are completely empty of population and businesses within 7-10 days, but it is unclear how the NBM differentiates between these locales.
In an effort to disentangle empty blocks from completely empty blocks, 2009 Census population projections are utilized (ESRI, 2009) in combination with 2010 business point data (ESRI, 2010). Figure 4 illustrates a basic typology of block areas for the city of Dublin, Ohio (n = 491). Blocks that are colored grey are locales where the residential population is at least 1 (n = 332). Blocks that are colored light orange have a population of 0, but are locations where at least one business or shopping center is located (n = 61). Notice that many of these blocks are located immediately adjacent to major transportation corridors – a pattern typical of many suburban commercial districts in the United States. In support of this general finding are the graduated symbols on this map which indicate the number of employees at each of the businesses in Dublin. In other words, these symbols are visual surrogates for the ambient daytime population. That is, larger symbols represent more employees at a business. Finally, the dark orange areas are blocks where residential population is 0 and there are no businesses or shopping centers within its boundaries (n = 98). These locales represent completely empty blocks and largely correspond to agricultural areas, large public green spaces, water bodies, or some type of undeveloped land. As noted previously, providers were asked to report areas where broadband service was currently available or could be provided within 7-10 days. For Dublin, nearly 46% of the completely empty blocks are reported as having broadband availability in the National Broadband Map. In other words, the NBM is reporting that 45 Census blocks which are completely devoid of residential population, businesses or shopping have broadband service available.

Clearly, statistics like these are difficult to accept. Why are some completely empty blocks labeled as having broadband while others are not? While providers may be able to reach these eligible locations with broadband within 10 days, including them in the final tabulation as being served is questionable. Consider the empirical evidence. As of 2010, it is highly unlikely that broadband service existed in the completely empty blocks because there is no person or business to serve in these

---

locations. While there is no doubt that population distributions change in suburban areas, settlement patterns in Dublin Ohio are largely “set”, and the areas highlighted as empty blocks with the population projections for 2009 and business point data for 2010 are likely accurate. Therefore, when the NBM includes these blocks and labels them “served”, the geographic extent of broadband availability is being drastically overestimated for the United States. This statement may seem like hyperbole, but considering the extent to which broadband service availability is overstated in a smaller, suburban municipality like Dublin, Ohio, it is likely that tens of thousands (if not more) blocks in the United States are tagged as having broadband available, when in fact they do not. In other words, regions that “could” have broadband are much different than regions that “have” broadband in the U.S.

5. Policy Considerations and a Blueprint for Improving the NBM

Creating the National Broadband Map is clearly a monumental task. In addition to dealing with a multitude of state agencies, broadband providers and a mass of both operational and spatial information on provision, developing a viable and unbiased reporting framework for these data is challenging. The basic empirical analysis provided in this paper suggests that the first iteration of the NBM is suffering from significant spatial data uncertainty and is likely far too generous in its estimation of broadband availability in the United States.

5.1 Policy Considerations

The overestimation of broadband availability in the United States is disconcerting for several reasons. First, history suggests that the closing of a technological divide is, at best, temporary. For example, the most recent incarnation of the digital divide shares a common history with Universal Service (i.e. telephone access) and the information haves and have-nots (i.e. computer access). As commercial dial-up service to the Internet began to blossom in the early-mid 1990s (Downes and Greenstein, 1999), the federal government passed the first major overhaul to telecommunications policy
since 1934. The Telecommunications Act of 1996 (96 Act) paved the way for a more competitive and dynamic environment for telecommunications and information services (Section 706). As a result, when broadband technologies such as cable and digital subscriber lines (DSL) began to replace dial-up services for both residential and small business Internet, empirical analyses suggested that broadband access was quite good in major urban and suburban areas, but lagging in exurban, rural and remote communities (Strover et al., 2002; Grubesic, 2003; Prieger, 2003, Greenstein, 2005; Flamm and Chaudhuri, 2007). Further, much like the information have and have-nots, the digital divide was structured along demographic, economic and educational chasms (Hoffman and Novak, 1999; Norris, 2001; Gabel and Kwan, 2001; Grubesic and Murray, 2004; Prieger and Hu, 2008). Once again, the alarm was sounded in the late 1990s (NTIA, 1999) and continued into the early 2000s, urging policy makers and the private sector to facilitate a more equitable rollout of broadband technologies. As noted previously, by 2010, empirical analyses suggest that between 4.5% and 7.7% of Americans do not have access to broadband (FCC, 2010b). This is not an alarming statistic, nor does it suggest that accessibility to broadband in the United States is highly problematic. Again, while the empirical evidence presented in this paper suggests that broadband availability estimates may be inflated, as early as 2001, Compaine (2001, 334) declared the war against the digital divide would be “won”, noting that by the end of the decade, the S curve of innovation diffusion (i.e. broadband diffusion) would flatten out by “allowing the self-evident forces of declining cost, natural acculturation and growing availability to take their course”. Compaine seems to be right.

However, even with the continued advancement of broadband technologies and growing levels of adoption, questions of availability linger. Why else would the government allocate $7.2 billion dollars to improve provision and access throughout the United States? In part, it is due to broadband dynamics. Because the broadband divide is a moving target, the technological evolution of access platforms (cable, fiber, wireless, etc.) means that many small and medium-sized urban locales and rural and exurban
communities are always at least a generation behind larger metropolitan areas in terms of access technologies. In part, this is because technological innovation is tightly coupled with return on investment. Most new telecommunications technologies are first provisioned to major metropolitan areas. That said, it is likely that some areas, particularly rural and remote locations, will never be targets for broadband investment from the private sector in the first and second wave of technology rollouts. Worse, there are some locations that may never be “worth” investing in because of the expense associated with provision and the lack of a dense subscriber base that can afford (or is interested in) advanced telecommunications technologies.

Second, hand in hand with broadband dynamics is quality of service (QOS) issues related to broadband performance. Not surprisingly, upload and download speeds vary considerably between markets, particularly when heterogeneities in platforms and platform generations exist (FCC, 2010a). In areas where competition is high, empirical evidence suggests that average top advertised speeds are also high – particularly when compared to markets where competition is sparse (FCC, 2010a). Further, broadband providers understand that opportunities exist to exploit cost and performance structures in competitive markets (Grubesic et al., 2011), offering different bundles (e.g. phone, television and internet) with varying levels of bandwidth at different price points. In areas where competition is relatively weak, providers are under no pressure to provide the fastest connections or competitive pricing. As a result, understanding where competition exists and the types of platforms and upload/download speeds which are available may help explain adoption rates, quality of service and the need for enhancing broadband markets with public policy.

Finally, issues of information asymmetry also play a role. Information on broadband subscribers is privately held by service providers. Until the National Broadband Map, the FCC and NTIA never reported provision information at levels below the census tract due for fear of putting providers at a competitive disadvantage. Notably, these types of information asymmetries are a common feature in
unsuccessful municipal investments in fiber and wireless broadband systems (Gillett et al., 2003; Gibbons and Ruth, 2006; Sirbu et al., 2006; Lehr et al., 2006). The abandoned WiFi network in Philadelphia, PA is a notable example. Built in conjunction with Earthlink at a cost of $20 million, it was immediately outdated, had coverage issues and was abandoned after only 6,000 customers signed up. Recent efforts by the city to resuscitate the network for use with public safety systems, government communications and targeted public access will require the city to spend nearly $17 million between 2011 and 2015 (Albanesius, 2009). Clearly, a much deeper and complete understanding of broadband alternatives is critical for regional economic development (Mack et al., 2011), social and political participation in the digital age (Sylvester and McGlynn, 2010) and the ability to conduct e-science (Andronico et al., 2011).

5.2 A Blueprint for Improving the NBM

Given the data uncertainty associated with the National Broadband Map, there are a few avenues for improvement that could be made in subsequent iterations of the NBM. As briefly noted in the introduction, the development of a robust and repeatable aggregation strategy for the broadband provision data is critical. Because the decennial Census information is limited at the block level, public policy analysts will be forced to aggregate block-level statistics to block groups or tracts for analysis. In its current form, the NBM data are extremely difficult to work with. Rather than having a single column that provides a total, unique count of broadband providers in each block, the NBM reports each instance of broadband provision and the associated provider. As noted in the case of large blocks (> 2 square miles), this means that 50 to 100 or more records may exist for each locale. While the web interface for the NBM does help sort out the mess of provision data for individual locations when queried, it would be helpful if the raw NBM data contained a summary column, at the block level, that distills this information into an easily accessible (and mappable) format.
Along these same lines, the NBM could summarize the number of unique providers within different administrative units, such as block groups, census tracts and counties. Again, these administrative units are critical to advancing public broadband policy because much of the most useful census and economic data are available for these geographies. This would greatly enhance the speed and ease in which analysts could develop econometric models of broadband provision and availability, facilitate the spatial analysis of broadband at multiple spatial scales and enhance the overall richness of the dataset by making it more easily accessible to local planning agencies and economic development organizations.

The National Broadband Map needs to be updated to reflect Census 2010 geographies. Rather than forcing analysts to use a clumsy and difficult Census block conversion table, it would be helpful if all future results published by the NBM used Census 2010 blocks. This is important for several reasons. First, there has been a dramatic increase in the number of blocks between 2000 and 2010. Again, Census 2000 included 8,262,363 blocks for the fifty states, Puerto Rico and the District of Columbia. Block totals increase 35% between 2000 and 2010, with a new total of 11,155,486 (Census, 2010). The driving factor for this increase is the improvement and enhancement of features included in the Census Master Address File (MAF) and the Topologically Integrate Geographic Encoding and Referencing (TIGER) data used by the Census Bureau, including tens of thousands of new roads and hydrographic (water) features. For example, roads within military and national park areas are no longer suppressed. Further, if the NTIA and FCC are truly committed to maintaining a current and realistic snapshot of broadband availability for the United States, the use of any data from 2000 hinders these efforts.

One final component of the NBM that could be improved is the provision of all NBM data in a completely disaggregated shapefile format or some other type of commonly used geospatial vector data file. The Environmental Systems Research Institute (ESRI) shapefile is a vector storage format for spatial data that contains points, lines and polygons for representing geographic information. When combined
with summary attribute tables from the Census Bureau and the broadband information available through the NBM, analysts would have access to a significantly enhanced geospatial database of broadband markets for the United States. Ironically, these data exist in shapefile format at the NTIA and are used to generate the gallery of maps available on the public site for the National Broadband Map (http://www.broadbandmap.gov/technology). However, the geographic files provided to the general public for analysis are provided in an aggregate format that does not neatly correspond to the native census block data used for NBM reports. For example, consider Figure 5a, which displays a portion of the Columbus, Ohio MSA, focusing on Franklin County. The areas shaded in blue represent the publicly available NBM shapefile for blocks less than two square miles for the state of Ohio.\(^{13}\) Clearly, these geographic coverages do not look anything like the native census block tessellation (highlighted in red) in Figure 5b. Again, it appears that these aggregate coverages were used to assign values to the underlying block tessellation. While issues of boundary uncertainty are a concern here (Grubesic and Murray, 2005), it would be much easier if the data were made available in a disaggregate shapefile format (i.e. native blocks) for download and analysis. Further, although wireless broadband availability was not a focus in this paper, it is important to note that wireless carriers actually provided maps of their coverage areas in shapefile format to the NTIA for the NBM – which were subsequently overlayed on blocks to determine the proportion of each block that is covered by wireless services (NTIA, 2011b). It is likely that providers are concerned that the full disclosure of the wireless coverage areas in shapefile format would put them at a competitive disadvantage with other carriers. For this reason, and the willingness of federal agencies like the NTIA and FCC to acquiesce to these requests, one should not expect these fine-grained data to ever see the light of day.

\(^{13}\) The file is available for download from http://www.broadbandmap.gov/data-download and is labeled OH-NBM-SHP-December-2010.zip.
6. Conclusion

In sum, efforts behind the National Broadband Map should be applauded. The United States is in desperate need of a comprehensive inventory of broadband availability. In its current form, although the NBM is far from perfect, it is a good start. From an analytical perspective, perhaps the biggest problem with the NBM is that the information collected is somewhat incomplete. As noted previously, provider participation varied significantly between states. Given these variations, one avenue for future research is to evaluate how sample selection bias may impact the evaluation of broadband coverage. Again, while the NTIA maintains that 95% of the overall market reported (Neville, 2011), the raw participation rates suggest that gaps in knowledge remain. For example, the empirical analysis in this paper used Ohio (83% provider participation). Thus, it would be interesting to compare and contrast results from other states where provider participation rates were somewhat different (e.g. Virginia [83%] and Indiana [100%]). Regardless, portions of the broadband landscape remains obscured. Thus, if the United States is to make headway on filling in these gaps and developing a better understanding of broadband availability and speed, there clearly need to be better incentives to provide the required information to the NTIA and FCC regarding broadband provision.

A final aspect of the NBM worth noting is the lack of reported pricing data. There is no legitimate excuse for not including this information. Again, it is likely that providers cite fears of “competitive disadvantage” as the reason for not reporting these data. In reality, it is likely that providers do not want consumers to see that subscribers in Philadelphia are paying 25% less than those in Cleveland for an identical package of services. Obviously, this is impossible to know because pricing data remains locked down by providers, but marked price differentials between geographic regions are likely, particularly between when locales with more than one provider are compared against those without any measure of competition.
References


IDInsight. (2011). Verification Analysis of the National Broadband Map. URL: [http://tinyurl.com/3opr85t](http://tinyurl.com/3opr85t)

Kolko, J. (2010). Does Broadband Boost Local Economic Development? Public Policy Institute of California. URL: [http://tinyurl.com/3frps8f](http://tinyurl.com/3frps8f)


National Telecommunications and Information Administration [NTIA]. (1999). Falling Through the Net: Defining the Digital Divide. URL: [http://tinyurl.com/3n8esdw](http://tinyurl.com/3n8esdw)


National Telecommunications and Information Administration [NTIA]. (2011b). National Broadband Map: Data Comparison Methodology. URL: [http://tinyurl.com/3sjv37u](http://tinyurl.com/3sjv37u)


Table 1: Example Records from a "Large" Block in Ohio

<table>
<thead>
<tr>
<th>Datasource</th>
<th>Frn</th>
<th>Provname</th>
<th>Hoconame</th>
<th>Fullfipsid</th>
<th>Transtech</th>
<th>Maxaddown</th>
<th>Maxadup</th>
<th>Downloadsp</th>
<th>Uploadspee</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROADSEGMENT</td>
<td>13430244</td>
<td>TIME WARNER CABLE</td>
<td>Time Warner Cable Inc.</td>
<td>390490080001000</td>
<td>41</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>13430244</td>
<td>TIME WARNER CABLE</td>
<td>Time Warner Cable Inc.</td>
<td>390490080001000</td>
<td>41</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>2946986</td>
<td>The Ohio Bell Telephone Company</td>
<td>AT&amp;T Inc.</td>
<td>390490080001000</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>2946986</td>
<td>The Ohio Bell Telephone Company</td>
<td>AT&amp;T Inc.</td>
<td>390490080001000</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>2946986</td>
<td>The Ohio Bell Telephone Company</td>
<td>AT&amp;T Inc.</td>
<td>390490080001000</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>2946986</td>
<td>The Ohio Bell Telephone Company</td>
<td>AT&amp;T Inc.</td>
<td>390490080001000</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>13430244</td>
<td>TIME WARNER CABLE</td>
<td>Time Warner Cable Inc.</td>
<td>390490080001000</td>
<td>41</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>2946986</td>
<td>The Ohio Bell Telephone Company</td>
<td>AT&amp;T Inc.</td>
<td>390490080001000</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>13430244</td>
<td>TIME WARNER CABLE</td>
<td>Time Warner Cable Inc.</td>
<td>390490080001000</td>
<td>41</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>ROADSEGMENT</td>
<td>13430244</td>
<td>TIME WARNER CABLE</td>
<td>Time Warner Cable Inc.</td>
<td>390490080001000</td>
<td>41</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

DATASOURCE - Source of availability data (Originally submitted to NTIA in RoadSegment or AddressPoint)
FRN - FCC Registration Number
PROVNAME - Provider Name
DBANAME - Doing Business As Name
HOCONAME - Holding Company Name (assembled by FCC)
FULLFIPSID - 2000 US Census Block 15 Character number
TRANSTECH - Technology Code (See below for valid values)
MAXADOWN - Maximum Advertised Download Speed (see below for valid values) from record level
MAXADUP - Maximum Advertised Upload Speed (see below for valid values) from record level
DOWNLOADSP - Maximum Advertised Download Speed if provided from Overview table
UPLOADSPEED - Maximum Advertised Upload Speed if provided from Overview table
Figure 1: Differences in Census Block Morphologies for Urban and Suburban Areas
Figure 2: xDSL Coverage for Central Ohio
Figure 3: Likely Service Overestimates for ADSL Provision in Central Ohio
Figure 5: NBM Provision Shapefile and Native Census Block Tessellation

a) Shapefile Reporting Broadband Provision for Census Blocks < 2 square miles

b) Native Census Block Tessellation

- Census Blocks
- Broadband Provision