# Measurement and Assessment of Broadband Availability

Report prepared for

John Adams Innovation Institute at the Massachusetts Technology Collaborative

By

William Lehr[1]
Tony Smith-Grieco

Massachusetts Institute of Technology

January 2009

[1] Corresponding author: wlehr@mit.edu. This report was prepared under the direction of Dr. Lehr, in collaboration with MIT graduate student, Tony Smith-Grieco. This work was funded by a research grant from the Massachusetts Technology Collaborative to support a joint MTC-MIT research effort to establish better metrics for broadband in Massachusetts, with support from the MIT Communications Futures Program (http://cfp.mit.edu), and the National Science Foundation. The opinions expressed in this report are of the authors alone.

## Table of Contents

# Executive Summary

With Governor Deval Partrick's leadership, Massachusetts has already approved legislation to create a $40 million broadband fund[2] and legislation is currently under consideration by the U.S. Congress to approve $6 to $9 Billion in Federal broadband funding as part of an omnibus stimulus package.[3] This is reflective of the recognition that broadband is essential basic infrastructure for our modern information-centric economy. As such, there is an enduring public interest in ensuring that all citizens and businesses have access to appropriate levels of broadband.

Given its importance and the commitment of significant funding to ensure universal access to broadband and healthy IT infrastructure, better data on the status and health of broadband infrastructure and service markets is needed to facilitate the efficient targeting of development funds, to evaluate the impact of broadband stimulus efforts, and to monitor the health of broadband across the state and across time.

This report builds on the earlier *Broadband Metrics Best Practices: Review and Assessment* (February 2008)[4] which surveyed current international, federal, and state efforts to assess the status of broadband service markets. The earlier study concluded that existing data was wholly inadequate to assess broadband availability and to provide a data-informed basis for formulating effective broadband policies, including targeting funding where needed. Rectifying this deficit will require significant resources.

The present report presents the results of follow-on research directed toward defining a practical framework and set of metrics to be used to identify and evaluate areas that are "underserved" (as opposed to "unserved") by mass-market broadband services. This reflects the recognition that an appropriate assessment of broadband represents a continuum of service levels and must consider multiple dimensions of availability, including:

- Service availability: where are services available?
- Choice and competition: can consumers choose among multiple service providers?
- Quality: what features (data rates) are supported?
- Price/affordability: are broadband services affordable to consumers?

In working toward a set of recommendations for how to assess broadband in Massachusetts, we focused on developing a practical proposal that reflected the fact that resources are scarce, data

[2] Governor Patrick signed the Massachusetts Broadband Incentive Fund bill into law on August 4, 2008. The bill provides funding in the amount of $40 million to provide high-speed broadband Internet access to underserved communities in Massachusetts. This bill also creates the Massachusetts Broadband Institute as part of the Massachusetts Technology Collaborative. The funds represent a one-time allocation to be expended over 36-months. (See Broncaccio, Diane (2008), "Broadband bill sets up 3-year window," Greenfield Recorder.com, August 6, 2008, available at: http://www.mtpc.org/broadband/news/GreenfieldRecorder8608.pdf).

[3] See Mark, Roy (2009), "Broadband, Net Neutrality Ride with House Stimulus Package," Government IT eWeek.com, available at: http://www.eweek.com/c/a/Government-IT/Broadband-Net-Neutrality-Ride-With-House-Stimulus-Package/.

[4] Available from: http://www.masstech.org/broadband/docs/BroadbandMetricsBestPracticesSurveyFeb2008.pdf.

constraints are endemic, and the best metrics are context dependent (e.g., vary with the question that is being asked). While more granular data offers greater flexibility in how the data may be presented and the questions that may be addressed, it also requires more resources, takes longer to collect, and may confront greater confidentiality/privacy concerns.

On the basis of our earlier analysis, and reconfirmed by our subsequent research, we believe that the base method for assessing broadband availability in the State needs to depend on the creation and management of detailed GIS-based maps similar to those that track the location of roads, power distribution, and water/sewage lines. This is consistent with the recognition that broadband services are basic infrastructure. The service providers are the best source of data for the creation of such maps.

In addition to developing detailed GIS-maps documenting the location (to within a few hundred feet) of broadband infrastructure (by service provider and by type of technology), a best-practices broadband metrics plan will need to undertake statistically valid (large sample) end-user surveys to collect data on adoption rates, end-user perceptions, broadband expenditures, and usage characteristics.

Both the creation of the detailed availability maps and the large sample end-user surveys, while necessary, are not the only data that should be evaluated in the context of a best practices approach. Additional data is needed for cross-validation and to address data gaps (including the need for current data while awaiting updates). Because the creation of the detailed GIS maps and large-scale end-user surveys were beyond the scope/resources of this project, we focused on identifying alternative sources of data and on defining a measurement framework that would be sufficiently flexible to be useful before the more detailed data is available and to incorporate the better data as it becomes available.

In the early phase of the project we proposed to develop a Broadband Community Index (BCI) that would allow one to develop a scalar score for a community (0 to 100) to enable communities to be ranked relative to one another in terms of their level of broadband service. The notion of a BCI proved conceptually useful and helped highlight recognition of the multidimensional and continuous character of broadband service availability. However, the BCI approach raises a number of practical challenges that resulted in us rejecting it at this stage in the evolution of broadband service metrics and markets. Implementing a single BCI condenses too much information and is too dependent on the choice of weights used to composite the individual metrics that comprise the BCI. Furthermore, comparing communities on the basis of a fine-grained scalar score suggests a level of accuracy that is not in keeping with available data and would likely be overly contentious.

Instead, we recommend that Massachusetts adopt a Broadband Report Card approach. This would retain the conceptual simplicity of the BCI approach (i.e., a limited number of metrics to consider, yet responsive to the multidimensional character of broadband), while offering greater transparency and flexibility without being as dependent on an arbitrary set of weights. Under this approach, a small set of "grades" would be scored for each community based initially on the answers to a series of questions that may be answered Yes/No/Unknown. (Later, as it becomes available, more detailed scalar data may be incorporated if further discrimination is needed.) The "grades" would be relatively coarse (reflecting the inherent uncertainty associated with ranking communities by their level of broadband service). We recommend adoption of a color coding scheme: “Green” = “Acceptable,” “Yellow” = Potential issue,” and “Red” = “Unacceptable.” As with a student’s report card, the collection of “grades” would allow one to gain an overall impression of broadband performance but may not admit to a strict ranking of communities. Thus, a community which scores “Green” in all categories is obviously better served by broadband than a community which scores “Red” in some categories. However, it may be less clear how to compare communities with mixed grades and, for some questions/contexts, certain grades may be more important than others.

We do not believe there is any single best version of a report card. The best metrics to include are likely to be driven by data availability constraints. Herein we discuss many of the issues and trade-offs that should be considered in developing appropriate broadband metrics. This discussion is intended to provide a toolset that will be useful to the analysts who will be tasked with the detailed work of designing and implementing broadband metrics for the state. Although we do not believe we have identified the best set of metrics, we do propose a version of an initial broadband report card (see exhibit below) that should provide an illustrative and useful starting point.

**Broadband Availability Report Card**

| Goal | Title | Description |
|---|---|---|
| | | ***Achieving Ubiquitous Availability*** |
| L0 | Town government is on-line | • Q01: Does the town hall have BB available for its internal use?<br>• Q02: Is there a municipal intranet linking (most) government offices, libraries, schools, hospitals, police/fire?<br>• Q03: Does the town maintain an active municipal website? |
| L1 | Public access BB is available | • Q11: There are 1 or more public access terminals with BB access in the community?<br>• Q12: Public schools have BB access for educational (and internal) use?<br>• Q13: Public libraries have public access terminals with BB access? |
| L2 | Ubiquitous BB available | • Q21: Are 1B broadband or better (“1B+”) services available to 95% or more of the homes? (Basic availability)<br>• Q22: Are 1B+ available from more than one facilities-based provider to 60% or more of the homes? (Competition and consumer choice)<br>• Q23: Are 2B+ available to 60% of homes? (Quality of service). This metric may be expanded to include other measures of service quality such as customer complaints or outages.<br>• Q24: Is price for 1B service at or below state average? |
| | | ***Keeping Broadband on track*** |

| | | |
|---|---|---|
| L3 | BB adoption on track | • BB adoption rates are on par with national average.<br>• $/Mbps/month for average, best, and entry service on par with national averages<br>• Within state differences on par with peer states. |
| L4 | BB is best in class | • BB availability and adoption rates for higher quality BB services (2B, 3B, 4B) are on par with national averages.<br>• Within state differences on par with peer states |

The motivation for adopting this design, the definition of terms (e.g., 1B, 2B), the scoring and options for modifying this report card are discussed in detail in Chapter 4 and elsewhere in this report.

# 1. Introduction

This report is the second of two reports prepared for the John Adams Innovation Institute ("JAII") of the Massachusetts Technology Collaborative ("MTC")[5] in order to assess and evaluate approaches for measuring the status of broadband Internet access ("BB") services in Massachusetts.

The first report provided an assessment of the current status and best practices experiences of efforts to collect, share, and analyze data on broadband Internet access services for mass market consumers in the United States and abroad (see *Broadband Metrics Best Practices Report*[6]). The earlier report identified the challenges confronting policy-makers seeking to frame and target effective broadband policies by the lack of adequate data on the current status and growth trajectory for broadband access services.

This report builds on the earlier report to provide guidance to the Massachusetts Broadband Institute, and others engaged in similar activities elsewhere on how to develop a broadband metrics framework that will be practically implementable, yet sufficiently robust and flexible to address the needs of policymakers and market participants (consumers, providers) for better data to track the status of evolving broadband markets.

This report discusses the challenges that must be confronted by any broadband metrics effort, sometimes with more detail than many readers may desire. We hope that this detail will be of most use to those engaged in the nitty-gritty details of designing and implementing a broadband metrics program. As we make clear, implementing appropriate broadband metrics will have to balance the potentially conflicting objectives of accuracy, cost minimization, timeliness, and understandability with pervasive data constraints. Selecting the optimal balance will depend on the context (what is the question to be answered? what resources are available to answer it?) and so it is not possible to recommend a single set of "best metrics" to use in Massachusetts. Instead, we provide a framework and discuss some of the challenges and trade-offs of adopting alternative proposals to provide a toolset with which to make the context dependent decisions required to keep a broadband metrics effort on track. Finally, for those who seek a more definitive recommendation about what to do, we offer an illustrative example of how the framework we set forth may be applied in Massachusetts to score communities relative to each other with respect to the status of their broadband service availability.

In the balance of this introduction, we first summarize the core lessons/recommendations that emerged from our earlier report and are reaffirmed by what we have learned in completing the present report. We then summarize the core high-level lessons learned in preparing this report, and conclude by briefly describing how the balance of this report is organized.

---

[5] See http://www.mtpc.org/institute/index.htm.

[6] See Lehr, William, Tony Smith-Grieco, and Grace Rusi Woo (2008), "Broadband Metrics Best Practices: Review and Assessment," report prepared for John Adams Innovation Institute of the Massachusetts Technology Collaborative (JAII-MTC), Februrary 2008 ("*Broadband Metrics Best Practices Report*", available at: http://www.mtpc.org/broadband/docs/BroadbandMetricsBestPracticesSurveyFeb2008.pdf).

## 1.1. Lessons from Benchmarking Broadband Metric Best Practices

In our earlier report, we summarized some important lessons learned from evaluating what others were doing with respect to assessing the status of broadband at the state, federal, and international level. The key insights included:

- ***Broadband is basic infrastructure***. As such, there is an enduring public interest in ensuring all citizens and businesses have appropriate access. Pursuit of this interest depends on market participants and policymakers having adequate data to assess broadband availability and needs. Moreover, data is needed for all classes of users/uses (not just those who may currently be least well-served). We take it as a given that just as it is inconceivable that public officials would not know where roads or electric service are inadequate, it is inconceivable that similar levels of detail would not be available to understand the status of broadband. A core characteristic of basic infrastructure is that it is something that is taken for granted until it is lacking.
- ***Data on broadband availability is inadequate.*** This is true at the international, federal, and state level. While Massachusetts is better than most, it does not have best-in-class data and even among those states with such data, there is significant room for improvement.[7] We do not collectively have the data we need to target broadband stimulus programs, assess the health of broadband service markets, and to understand the larger implications (economic impacts) of broadband on our economy and society.
- ***Broadband metrics presents a difficult challenge***. The problem of collecting adequate data is complicated by the fact that broadband is still a relatively new mass-market service, and the markets/technology are continuing to evolve rapidly. In contrast to other forms of infrastructure that are publicly owned or provided by a single monopoly provider (e.g., roads, local water and power distribution), broadband services are typically provided by investor-owned service providers in increasingly competitive retail service markets. Finally, broadband is inherently multidimensional. Appropriate metrics need to consider technical characteristics (e.g., what platform? What peak/average data rate downstream/upstream?), price, and consumer choice attributes (e.g., extent of competition?).
- ***The "broadband metrics" problem is accentuated by the lack of consensus regarding the appropriate definition of what constitutes adequate broadband service.*** While it is common for broadband stimulus proposals to discuss the need to target "areas that are unserved or underserved by broadband" there is a noted lack of guidance regarding the specifics of how to identify the appropriate "areas" (e.g., what geographic granularity is appropriate for measurement?) or develop rankings to determine when one area is worse off than another with respect to broadband service (e.g., what service attributes should be included in the assessment and how should they be weighted?).

[7] California, Kentucky, and North Carolina are generally recognized for their early progress and success in mapping the availability of broadband. However, even in these states, there are questions about how often the data will be updated and regarding the accessibility to the data. This limits analysts' abilities to integrate the data with other third-party data (e.g., on demographics).

- ***GIS-based, statewide mapping of broadband infrastructure is needed.*** The best-in-class states have implemented Geographic Information System (GIS)-based maps based on data collected directly from service providers that shows where those service providers have broadband services.[8] Ultimately, these maps and the data on which they are based will provide the best data on broadband infrastructure and service availability. These maps are interactive and capable of being integrated with other data layers showing standard mapping attributes like town/county boundaries, roads, and other topological features at varying degrees of granularity (within a few hundred feet).

## 1.2. Goals and Lessons from Measuring Broadband Availability

This report extends the earlier analysis by examining the broadband metrics challenge more closely. The research on which this report is based was conducted during the first half of 2008, in collaboration with personnel from MTC-JAII and other experts across the State. The goal was to define a *Broadband Community Index ("BCI")* that could be practically implemented to allow one to rank communities based on their relative level of broadband service. We decided to focus on the availability of mass market broadband access services, rather than on access services to medium or large enterprises, or other elements of broadband-related infrastructure (e.g., back-haul facilities, data centers, or switching facilities). Moreover, we proposed to test the viability of the BCI approach using a mix of data collection strategies that are alternative to and complementary to approaches based on collecting the data directly from service providers (which is the preferred approach for building the sorts of detailed GIS maps described above). Finally, we proposed to test the practicality of the proposed methodology by applying it to a sub-sample of the 351 towns/communities that comprise Massachusetts. While the focus was on Massachusetts, the analysis applies more generally. This report presents the results and lessons learned from this research effort.

Several key conclusions emerge from our analysis:

- ***Core conclusions from the first report were reaffirmed.*** Importantly, we concluded that a statewide GIS-mapping effort is necessary and that service providers will be key sources of the necessary data. Undertaking this effort will require a commitment of significant resources (on the order of $500-$1million or more with on-going annual expenditures on the order of $50-$100k to keep these maps up to date).[9] These maps should provide a comprehensive view of the geographic location of key IT infrastructure (e.g., broadband access facilities, long-haul fiber, switching centers, cable head ends, antenna sites, etc.) and be comparable in detail to the sorts of maps used by road planners or those tracking other infrastructure such as electric power distribution or water and

[8] As noteworthy examples, see California (http://www.calink.ca.gov/taskforce/appendix_maps.asp) and North Carolina (http://www.e-nc.org/disclaimer.asp). While the detail reflected in these maps is noteworthy, there are obvious limitations to data accessibility. For example, in the case of California, the maps are provided in PDF format which makes it difficult to integrate the data with other sources. Also, the way speeds are classified may give a misleading impression of what is available to the average consumer.

[9] These are ballpark estimates. The cost of acquiring and maintaining broadband and related IT infrastructure data will decline as standards emerge and experience accumulates. There ought to be realizable scale economies from sharing best-practice experiences.

sewage. Further elaboration of what is required to design and implement such maps, while necessary, is outside the scope of this project. Consequently, we focused on alternative data sources (i.e., not provided directly by the service providers to our research project).

- ***GIS-maps, while necessary, are not sufficient.*** While we concluded that the GIS-maps of the sort described above represent a necessary element in any best-practices approach for measuring broadband, this is not a sufficient source of data and other approaches for data collection are available and necessary. These include web-based searches, end-user surveys, and potentially, “expert” questionnaires. Such additional approaches are needed to supplement the map-based data (e.g., to learn about consumer opinions or to collect data on service pricing); to cross-validate the map data; and to answer questions when current GIS-map data is not available.

- ***Broadband metrics face enduring data constraints, so the best metrics are opportunistic.*** Even if one ignores the challenges of designing and constructing appropriate GIS broadband maps, the alternative data sources all face difficult challenges to implement cost-effectively. First, there is the challenge that expert knowledge about the status of broadband services is scarce outside of the service providers and/or expensive to acquire (e.g., professional consultants). Surveys of end-users, while useful and necessary to assess perceptions and some aspects of usage or pricing characteristics, are much less appropriate for measuring service availability. Most consumers are not aware of what services are available. Moreover, early on, we anticipated being able to implement an “expert questionnaire” of IT town managers or other local “volunteer experts” on the local status of broadband across the state. This was rejected when we considered more closely the practical implementation challenges (e.g., identifying a cohort of suitable experts who could complete the proposed questionnaire cost-effectively). Second, indirect sources of data such as from web-based research (e.g., by pinging service provider websites) or third-party market research is either prone to variable quality (e.g., websites change and provide inconsistent and insufficient details) or expensive (e.g., market research firm data). Third, because broadband markets are evolving quickly, the currency of data is short-lived. This means that the best broadband metrics are opportunistic: make the best use of what you have available. Analysts need to be flexible and appreciate that broadband metrics will evolve – not just because broadband evolves, but because available data on broadband will evolve also. In the following, we will discuss the pros/cons of a variety of approaches, and will make some recommendations, but the best approach in a specific context will depend on the context. The recommendations made here are, in many cases, illustrative rather than definitive. Our goal in presenting them is to suggest one way that broadband service might be scored. Understanding the tradeoffs from using alternative data sources/metrics is important to make the best use of available data.

- ***Report cards are better than a Broadband Community Index.*** While it is conceptually useful to imagine developing a scalar BCI that provides a numeric score for a community to measure its level of broadband service, there are practical reasons why doing so is problematic and not recommended. First, the concept of such a scalar index is useful because it suggests the ability to rank-order communities into those that are better or worse served. This is important because it forces recognition that the level of broadband service availability is a continuum. For example, there are few environments when broadband is wholly unavailable, and yet many more where the level of broadband availability reasonably may be regarded as inadequate. Moreover, it also correctly reminds folks that an index that weights multiple metrics may better capture the multidimensional nature of broadband service than any single measure such as the peak data rate or share of households that may be able to subscribe to broadband services in any given area. Finally, a BCI is attractive because it provides an intuitively obvious way to simplify data analysis. However, implementing any specific BCI is practically problematic because the weights used to comprise the index are inherently arbitrary and best chosen with an eye toward the specific use or question to which the index is being used.[10] Moreover, a numeric score suggests greater precision than is warranted given likely data constraints and disagreements over how weights ought to be assigned. The challenge is akin to the challenge of scoring student performance with a single numeric grade. Analogously, instead of offering a strict ranking methodology, we propose a report card approach that makes use of a set of relatively simple and coarse color coded metrics ("green, yellow, red" for "good, potential issue, inadequate"). The ambiguity of colors instead of a strict numeric score highlights the inherent uncertainty in measurements and is simpler to implement. At a minimum, the report card should focus on three aspects of broadband availability: (a) the extent of broadband coverage; (b) the price/affordability of broadband; and (c) competitive alternatives. The framework is sufficiently flexible to allow one to add additional "grades" and admit diverse weighting criteria.

- ***Broadband services are local and assessment ultimately needs to focus on the individual household.*** In undertaking this analysis, we focused on Massachusetts' towns/communities as the initial entities for analysis. While these jurisdictional boundaries offer one obvious way to report the data, they are arbitrary and do not match the contours of broadband service markets. For end-users, it matters what services are available at their household, not what services are available somewhere a mile or more away in their town.[11] Practicality constraints are likely to dictate the granularity of data available. In some contexts, it may be available only at the county level, the zip code level, the Census Block Group, or the town or municipality level. More granular data may be aggregated into larger geographic regions, but the reverse is more problematic.[12] Ultimately, we would like to know what is available at each household or point on the map to assess the level of broadband service in any particular area of interest (which may coincide with but need not be limited to town boundaries). Focusing on the point data and allowing variable geographic area aggregations makes the approach more flexible and robust, and may help immunize the broadband metrics process from political squabbling among communities.

[10] For example, BCI may be different if the goal is to compare average level of service across communities as opposed to identifying outliers.

[11] For example, the quality of available DSL services declines with the distance from the central office when loop lengths are long. Service may not be available or available only with greatly reduced peak data rates to households that are relatively far from the central office.

[12] For example, knowing that broadband service is available in part of a zip code does not allow one to know where in the zip code it is available.

### 1.3. Report organization and guide to readers

The balance of this report is organized into four chapters, and a series of appendices. Chapters 2 and 3 provide a general review and discussion of the broadband metrics problem. In Chapter 2, we propose a definition of broadband and identify the classes of data that should be included in the broadband assessment and some of the criteria that should be used to identify appropriate measurement strategies. In Chapter 3, we consider the various types of metrics, the issues associated with constructing/using a BCI, and examples of some of the types of individual measurements that may be appropriate.

In Chapter 4 we present one version of a potential "report card" to illustrate how the level of broadband services might be assessed. Our proposal/example incorporates the key lessons that emerge from considering the metrics challenge in greater depth. While we believe it offers a practical solution as presented, our expectation is that it will provide a useful starting point for an on-going discussion among those engaged in the measurement effort.

We conclude in Chapter 5 by presenting our analyses of a variety of different types of publicly available data on broadband availability across Massachusetts that were identified by/collected by Tony Smith-Grieco during this research project. These data demonstrate both the limits and opportunities afforded by third-party data sources. Finally, the Appendices provide additional detail on the data sources, and other data collection issues.

## 2. Understanding the Broadband Metrics Challenge

One key data challenge is the lack of generally accepted standards and definitions for what constitutes "broadband" service. For example, while most analysts agree that the FCC's prior standard which defined broadband services as those offering a data rate in excess of 200Kbps was too slow, there is no general consensus as to what higher rate may be best.[13] The lack of such consensus hampers comparisons of such limited data as exists. Obviously, the more granular the data collected (e.g., in multiple speed tiers, by technology, and with other attributes such as pricing), the easier it is to re-aggregate the data to fit whatever definition one adopts. However, collecting more granular data is more expensive, potentially subject to greater measurement error, and is more likely to raise concerns about confidentiality. Thus, an appropriate balance must be identified to ensure that the data collected is appropriate to the questions it will be required to inform and the resource burdens implied in its collection, on-going maintenance, and access management.[14] Furthermore, because the market measurement and evaluation tasks will be on going, the mechanisms need to be dynamically sustainable (repeatable, robust to changing market conditions and technology, and inter-temporally comparable).

[13] Subsequently, the FCC has announced plans to upgrade its data collection efforts, but it is unclear how this data will be made available to State policymakers and the quality of the ultimate data remains unclear. See Federal Communications Commission press release, "FCC Expands, Improves Broadband Data Collection", March 19, 2008. More information at http://www.fcc.gov/headlines.html.

[14] For example, the more granular the data is, the more likely there will be confidentiality concerns, and consequently, the greater the likely costs will be for managing access and enabling integration with third-party data.

The *Broadband Metrics Best Practices Report* explained how defining what constitutes broadband service is complex and multi-dimensional (i.e., needs to address technical characteristics of the service,[15] competitive options, and pricing) and why simplistic binary classification schemas are problematic. For example, classifying communities as "served" or "unserved" obscures the fact that there are shades of service availability. It may be more meaningful to focus on defining an index to give a scalar measure of the level of broadband service in a community. Hypothetically, one could imagine constructing such a *Broadband Community Index* (BCI) as a weighted composite of a series of individual metrics referencing different service attributes. The index, the individual metrics, and the measurement strategy for each metric would collectively define a mechanical process by which one could produce a scalar index value for a community (or other region) that would allow communities to be ranked in terms of their levels of broadband service. Were such an index constructed, it would then be possible to develop a classification scheme that could be binary (e.g., communities with a score lower than X are identified as being *underserved*).[16] Intuitively, such an index might prove useful in targeting broadband promotion resources (e.g., public investment funds), in evaluating progress over time (including economic impacts), and in facilitating benchmarking within the state and across other jurisdictions.

Additionally, the *Broadband Metrics Best Practices Report* explained that the state of the art in broadband service measurement requires detailed GIS-based mapping.[17] Such an approach is needed to accurately ascertain the availability of broadband services at different locations in a community and options for consumer choice and for improving availability in the future (e.g., costs of expanding coverage). However, acquiring, maintaining, and presenting such data is expensive and will likely require the cooperation of service providers across the state. Some data is available indirectly in the form of strand maps from cable operators which are publicly available from the DTC[18] and information about the locations of ILEC COs with (and without) DSL-enabling DSLAMs, this data is less accurate than what the service providers have. While accessing third-party data is likely to prove useful for cross-validation purposes or in lieu of carrier-provided data (before such data may become available), we believe that the carriers ought to be the principal providers of data to the mapping effort. Figuring out how to induce the carriers to make such data available (e.g., whether voluntarily or in response to regulatory rules) and determining appropriate access rules to ensure adequate data protection pose challenges that need to be addressed.[19]

---

[15] For example, peak and average data rates in upstream and downstream directions.

[16] You could also have multiple classifications such as "poor," "average," and "best" based on alternative aggregation approaches.

[17] As identified in the earlier report, the detailed maps of broadband service availability have been prepared in a number of states with the cooperation of the service providers and include overlapping maps of the infrastructure from the various service platforms. Examples of states with such detailed GIS maps include California, Kentucky, Maine, North Carolina, Vermont, Tennessee, and Wyoming.

[18] These strand maps are of variable vintage and quality.

[19] Data confidentiality may be an issue for strategic/competitive, security, or end-user privacy reasons. Further discussion of these issues and why carriers may be reticent to share detailed data or how them might be incentivized to provide such data is beyond the scope of this project.

While detailed GIS-based mapping that shows the location of broadband infrastructure (access facilities, antenna sites, central offices, cable head ends, longhaul fiber, etc.) at granular resolutions (within a few hundred feet) is deemed to be an essential component of a best-practices measurement approach, it is not sufficient. Other mechanisms for collecting data are also needed to provide cross-validation, support analyses during the times when the GIS data is not yet available (or between updates), and to support more flexible "what-if" analyses and integration with data that are not part of the GIS.

One obvious alternative is to rely on community-based surveys to supplement the GIS data on broadband availability. On the plus side, such data may be collected relatively quickly and with little expense (especially if implemented on-line). Community-based surveys also may provide data beyond service availability that is of interest (e.g., regarding customer attitudes and tastes). Surveys may prove useful in evaluating adoption behavior (subscribership/service penetration), expenditures (service pricing), and usage (what applications are subcribers using? How intensively are they using broadband?). Additionally, community-based data collection may help raise broadband awareness and mobilize community engagement. Increased community engagement will help promote broadband adoption and provide a channel for identifying local opportunities and challenges on an on-going basis. However, community-based surveys are likely to do a poor job in assessing service availability. Consumers are notoriously poorly informed about options they do not have direct experience with. While we may believe they know about the service they use if they are subscribers, we should regard with caution their assessment regarding other options.

Another obvious alternative is to make use of online resources via the Web as a platform for mounting end-user surveys and data collection from public websites. For example, the service providers maintain retail websites where actual and potential broadband customers may check service availability or order broadband service. Additionally, there are a number of third-party mounted data collection efforts such as SpeedMatters.Org, a site supported by the Communications Workers of America (CWA) that provides an on-line test of broadband speeds.[20] Tapping into such indirect data sources provides another source of data to cross-check broadband availability and possibly other relevant data about service quality and competitive choices. For example, the provider site data offers an indirect way to access service provider data and could be used, in conjunction with end-user survey data to cross-check survey responses and thereby enhance survey accuracy. The broadband speed data provides some indication of service quality (tests of data rate variability) and availability (i.e., broadband is available where it is tested, however you do not know in areas without test results whether it is because broadband is not available or because no one has elected to take the test).

Offsetting the benefits of such non-GIS-based data collection efforts, are the fact that community-based surveys are likely to be less reliable[21] and face implementation challenges. For example, telephone or mail-based surveys are more expensive to mount than Web or email-based surveys. While less expensive, email/web-based approaches may make it more difficult (if not

[20] See http://www.speedmatters.org/. Also, see the discussion of Speedtest.net further below.

[21] Significantly improving the reliability of community survey data for identifying what/where broadband services are available would require additional resources that offset one of the key advantages of this approach.

infeasible) to reach end-users who do not have broadband service (i.e., either do not have service available, or if available, have elected not to adopt). Additionally, obtaining informed answers about a complex service such as broadband provides an additional challenge.

## 2.1. Defining "underserved"

Broadband availability is seldom an absolute. There are levels of availability: some locations have better access to broadband options than others. This suggests that it may be better to focus on defining a BCI that would allow communities to be ranked relative to each other and then classify those rankings into groupings that are better or worse served.[22] Instead of focusing on communities that are "unserved," one should focus on communities that are underserved, whereby the definition of what constitutes being underserved may depend on the context.

As part of this project we reviewed the broadband literature, grant criteria for a number of existing broadband subsidy programs, and interviewed a number of broadband experts regarding the current status of formal definitions that would allow one to unambiguously define whether a community is "unserved" or "underserved." None of the experts or sources considered defined "underserved" with respect to broadband capacity in any great detail. Examples include the USDA Rural Utilities Service Rural Broadband Access Loan (and related programs), the FCC Universal Service programs, the Michigan Broadband Development Authority and the Maine ConnectME Authority. Some of these sources require a community to have no broadband service whatsoever. Most of them are based on some demographic factors about the community, such as rurality, per capita income, or qualification for other subsidy programs (such as the national School Lunch program). There have also been various efforts to define more general "e-readiness" measures, for example by the *Economist* magazine[23] and the World Economic Forum.[24] However, in these cases either the detailed metrics are not published or the published metrics are very general.

Because there was a noted lack of formal definitions of what identifies the state of being underserved by broadband, we sought to develop our own metrics, building on the analysis in the *Broadband Metrics Best Practices Report*. This entails the following sequence of steps:

(1) Define broadband service

(2) Define scope for what constitutes access

(3) Identify criteria for selecting appropriate metrics

(4) Define metrics

(5) Develop methodologies for measuring the proposed metrics

---

[22] Because of measurement error and ambiguities in defining metrics, it may be preferable to group communities into coarse categories (e.g., quartiles) rather than reporting the numeric Broadband Community Index.

[23] http://www.eiu.com/site_info.asp?info_name=eiu_2007_e_readiness_rankings

[24] http://www.weforum.org/en/initiatives/gcp/Global%20Information%20Technology%20Report/index.htm

Each of these steps is described further below.

## 2.2. Defining "broadband"

In the *Broadband Metrics Best Practices Report*, a classification schema for broadband service was presented, based on the bandwidth provided: (p. 28)

**Defining Broadband**[25]

| Level | Data rate | Technology Platforms | Services enabled |
|---|---|---|---|
| *0B* | 50Kbps | Dial-up modem – not considered to be broadband. | Pre-broadband Internet access |
| *1B* | 500Kbps | 1st gen DSL/Cable modem service, 3G wireless, satellite | Email, web browsing, VoIP |
| ***2B*** | **5Mbps** | **2nd gen DSL/Cable modem, WiFi, WiMAX, EV-DO Rev A (2Mbps), HSDPA (3-7Mbps)** ***(what most mass market broadband customers have today)*** | **Streaming video, rich interactive media** |
| *3B* | 50Mbps | xDSL, FiOS FTTH, next-generation cable (DOCSIS 3.0) – WIMAX, LTE, 802.11n | Multichannel video, Triple play |
| *4B* | 500Mbps | Next gen FTTH/"λ access" | Telepresence |

We believe this schema represents a reasonable balancing of the need for a simple definition and the reality of comparing the complex array of differentiated service offerings available in the market.

### 2.2.1. Broadband data rates as a proxy for service quality

To enhance the usability of the above, we would suggest that the data rates identified above be specified as a range to allow more flexibility in classifying services. For example, 0B could correspond to services offering data rates of from 0 Kbps to 200Kbps; 1B to data rates from 201Kbps to 1.5Mbps; 2B from 1.6Mbps to 20Mbps; 3B from 21Mbps to 100Mbps; and 4B being services with greater than 100Mbps. The point estimates suggested above are meant to be illustrative and suggestive of the 10-fold data rate increase that contributes to defining a substantively different level of service worthy of a further refinement of the classification schema.

[25] p. 4 of Best Practices Report

Additionally, we have focused on the peak download rate since this is the single most easily observed rate. However, a more complex analysis ought to consider both upload and download rates. Upload rates are important for interactivity and are usually substantially lower than upload rates, sometimes by an order of magnitude. In California, they address this by summing up and download speeds. This strategy is interesting because it pays attention to upload speeds without requiring us to keep track of two numbers, but is problematic because it limits comparability with other data since few others who report broadband data rates aggregate in this way.

As we discussed earlier, even the focus on peak data rates is misleading since this may be poorly correlated with average data rates (up or down), and because peak and average data rates have different implications for different types of applications.

The best that can be said for focusing on a single rate is that it is relatively simple and when coupled with the identification of the technology platform provide a rough yet reasonable basis for assessing the quality of broadband service available.[26]

We are using data rate as a proxy for service quality but this is only approximate. Generally, higher data rate services are supported on newer, more advanced infrastructure platforms that also offer additional service enhancements. Also, higher data rate services can support a wider array of services (i.e., on a 50Mbps broadband channel you can do both email and stream high quality video; while on a 1Mbps broadband channel you can do email but cannot stream high quality video). However, other aspects of service quality are also important and may be unrelated to supported data rates. For example, the incidence of customer complaints or outages, the time-to-repair, and other technical parameters such as jitter, packet loss, or end-to-end delay may be relevant in assessing quality. A recent report by the UK regulator, Ofcom highlights this point and demonstrates how congestion (a byproduct of changing end-user demand patterns and service provider provisioning decisions) may impact realized data rates over the course of a day.[27]

---

[26] This is quality in terms of the offered service. The actual quality experienced by consumers may depend on network loadings and management practices. Monitoring such things as customer complaints or calls to customer service or customer satisfaction surveys may also provide additional information on the quality of broadband services.

[27] See Ofcom, "UK broadband speeds 2008: Consumer experience of broadband performance, initial findings," January 8, 2009 (see http://www.ofcom.org.uk/research/telecoms/reports/bbspeed_jan09/bbspeed_jan09.pdf). As part of this study, Ofcom relied on an infrastructure of edge-based measurement boxes that collected data on the following sorts of measurements:

- Latency (ICMP based)
- Packet loss (ICMP based)
- DNS query resolution time
- DNS query failure rate
- Web page loading time
- Web page failure rate
- VoIP call quality (MOS)
- VoIP call jitter, delay, packet loss
- SMTP relaying speed
- Web based download speed test (HTTP)
- Web based upload speed test (HTTP)

A richer assessment of broadband quality ought to expand beyond consideration of data rates to include the additional sorts of metrics mentioned above.

### 2.2.2. Which broadband platforms to include in assessment

At a minimum, the assessment of broadband ought to include DSL and cable modem services offered by telephone and coaxial cable television providers, FTTx services,[28] fixed broadband wireless (e.g., from a WISP using WiMAX or other wireless technology), and broadband-over-power lines. The last two technologies have not been as widely deployed but may offer more important vectors for competitive choice in the future and may prove especially interesting in addressing rural areas that have heretofore been poorly served by more traditional wired alternatives.

Excluded from the above definition are satellite and 3G mobile wireless services, although this assessment is likely to need to change over time. The rationale for excluding these types of wireless services is that they generally offer lower quality service[29] than is available from wired infrastructure and it is hard to verify the actual availability of wireless services on a per home basis. The satellite option is generically available everywhere, and whether it is actually available to an individual homeowner may depend on whether it is actively marketed in the region and whether there is a clear enough line-of-sight from the home to the satellite. Evidence that a significant number of users are subscribing to satellite services in an area or perceive it as a viable option would argue in favor of including it in an assessment of broadband. Improvements in satellite offerings and the introduction of anticipated mobile satellite broadband services may require further reassessment.

With respect to 3G services the challenge is more complex. To date, 3G services have been marketed as a mobile broadband complement rather than substitute for fixed broadband services. However, with the expansion of 3G infrastructure, improvements in service quality and price reductions, 3G may need to be considered in the assessment. If and when evidence accumulates that mass market consumers are viewing 3G as a viable substitute for wired broadband (analogous to the evidence that a significant number of consumers now view mobile telephony as a viable substitute for fixed POTS telephony[30]), it will be necessary to include 3G in the assessment of broadband availability.

---

- Non web based download speed test
- Non web based upload speed test
- Single stream download speed test
- Multi-stream download speed test
- Single stream upload speed test
- Multi-stream upload speed test

[28] This includes fiber to the home (FTTH), fiber to the curb (FTTC), and fiber to the building (FTTB) services.

[29] Lower quality in the sense that the data rates are usually lower for satellite and 3G services than are available from wired or fixed wireless broadband alternatives, especially in the upstream direction.

[30] The number of consumers who have only a mobile phone is now in the double digits and is quite high among certain classes of users (e.g., college students). The growth of mobile telephony and VoIP-over-broadband have contributed to significant reductions in the number of second lines.

Including 3G (and/or satellite broadband, including mobile satellite broadband) will force further consideration of the modalities of broadband use since a key feature of such services is that they support mobile access. Much of the future growth in broadband use and applications, and much of the excitement in such growth rests with mobile services that support ubiquitous/always available/always connected communications. The growth of WiFi and the nomadic access it supports and the growing market for 3G services suggests how adding mobility expands the usability of broadband and forces us to reconsider its essential characteristics. For now, however, the focus is on fixed line broadband access.

After defining what constitutes broadband for the purposes of measurement, it is necessary to define what constitutes "access to broadband." As we explain below, this requires identifying *who* has access, *where* they have access, and at what *price*, and what *competitive* choices are available.

### 2.2.3. Focus on residential, not business users

In this research we focus on mass market services. The basic rationale for this is to keep the project focused. We believe medium and large enterprises also face service concerns, but that assessing their concerns presents additional complications that need to be separately addressed. First, broadband service is more likely to be available for businesses than residential users, and it is rare for medium or large enterprises to be without any kind of broadband service – although that does *not* mean that the broadband that is available is adequate either in terms of price or quality.[31] Additionally, the requirements of businesses for broadband access are much more heterogeneous (e.g., variability in establishment size and data connection requirements across business activities varies more widely than for typical households). Thus, it is possible that areas with adequate residential service may lack adequate service for business (e.g., access to fiber); or visa versa.

### 2.2.4. Include public shared terminal and home access options

In addition to identifying the class of broadband users, it is important to consider the mode of access. Residential consumers may access broadband in a variety of venues, including from public access terminals (in libraries, town offices, schools, etc.), at work, and at home. We believe that ready access to public access terminals and home access reflect two important classes of availability that ought to be considered separately. We exclude access at work for the same reasons as why we limit consideration of broadband services for businesses.

While the principal focus of broadband availability is on home access to broadband services, this does not adequately reflect the full range of access options, especially at the low end. It is also important to focus on public (shared) access options. Moreover, focusing solely on whether broadband services pass all homes in a community does not mean that all residents have adequate access to the broadband Internet. For example, some residents may not have a home PC or the requisite expertise to use or funds to acquire one. Residents who are within reasonable proximity (say, a few miles) of a public access terminal in a library, government building (e.g.,

[31] Knowledge that service is available to residential users usually allows one to infer that service is available to small businesses also. Analogously, knowledge of commercial availability of services provides evidence that basic infrastructure is in place that, in principal at least, might be expanded to provide service to residential consumers.

town hall), or public school might be considered as having access to the broadband Internet, and certainly better at least than those whose only option is to travel much further for equivalent access. Public terminal access, while lacking the convenience of home access, offers several advantages. First, it is in-service which means accessible to broadband users who might not find it desirable to subscribe to home access services even if available.[32] Second, a public access terminal in a library has the requisite customer-premise-equipment (i.e., a PC with appropriate access software) and is accompanied by limited support (i.e., the librarian who can assist with access questions). And, third, such terminals provide points of access for on-line communications with the community (e.g., such terminals may be used to complete community-based on-line surveys). Since many of these public access terminals (in schools, public libraries, and/or government buildings) may be supported via expensive T-1 lines, the availability of such broadband – while valuable – does not provide evidence that adequate mass market broadband services are generally available to households in the community.

### 2.2.5. Affordability

The affordability of access is also relevant to consider. If the same services are available to the same share of homes in community A and B, but those services are priced much higher in A or residents in A are much poorer (so the price represents a higher percentage of their income), then one might argue that A is underserved relative to B. Such information is relevant to understand differences in adoption rates (what really matters for broadband to have a significant impact), private investment incentives, and the needs for public policy interventions. Because services vary in terms of peak data rates, the measurement of affordability should focus on a comparable metric such as the $/Mbps/month, and may need to be normalized to account for per capita income differences.

Because most carriers offer services over a wide area, and offer standardized broadband offerings (which may be national in scope), there is relatively little variability in broadband pricing (within a carrier). Such variability as exists is more likely to exist across carriers or across service offerings (e.g., not all service packages are available in all locales in the service providers footprint). Additional variability occurs as a result of periodic retail campaigns that may offer locally targeted discounts.

The relevant data on service pricing needed depends on the type of question you are asking. For example, if you are focusing on the average level of service in a community, you should focus on average rates (i.e., weighted by the services actually purchased by consumers); whereas if the focus is on assessing the minimum adoption costs for marginal residents then you may wish to focus on the lowest tier pricing offered. You may be able to infer the former from customer survey data, while the latter may be available by looking at advertised provider offerings. Note that these are two very different types of data and appropriate to addressing related, but different concerns. Generally, we prefer data based on actual expenditures since data based on offered services may be hard to evaluate. When data from advertised services is used, we recommend focusing on those services that are most widely adopted by the mass-market consumers. Offers which are rarely adopted may not be widely marketed (which means they may appear as

[32] For example, their demand may be intermittent or insufficient to warrant paying for 24/7 broadband access. Such consumers might prefer dial-up at home and access to broadband at a public access terminal for their more occasional special needs.

"unavailable" to consumers) or may have other features that make them less attractive than their advertised price may make them first appear.

### 2.2.6. Competitive Choice

Finally, one might reasonably be concerned with the number of facilities-based[33] competitors that offer competitive choices in a community. Communities where there are multiple service providers offering broadband service are better served by competition than communities with only a single provider. Increased competition is deemed beneficial because it increases the likelihood that consumers will have a range of differentiated service options to choose from and that they will benefit from the beneficial impact of competition on retail pricing and service quality. As already noted, the principal local benefit from more intense competition with respect to pricing may be realized in the form of an increase in discount offerings. To the extent competition results in lower pricing, these benefits are likely to spillover across communities as when provider A lowers its state-wide pricing to address the relevance of provider B's competitive threat in a subset of provider A's markets. With respect to service quality, the benefits from more intense competition may be harder to assess. For example, providers may be inclined to target investments in improving service quality to markets where consumers face more alternatives and the providers feel greater pressure to keep their subscribers happy.

## *2.3. Geographic Focus for Analysis*

Another important consideration is how to identify the appropriate regions for assessing broadband availability. As already noted, the best data would identify the state of broadband availability at every possible location (i.e., a detailed GIS-based map). What would be of most interest would be the availability of broadband for every household in the state.

Were sufficiently granular data available -- identifying broadband availability on a per-household basis -- it would be possible to aggregate such data up to larger geographic areas on the basis of jurisdictional boundaries (towns, counties, state), topographic features (proximity to major roads), economic development/integration areas (MSAs), coverage areas of specific providers, or some other basis (the geographic coverage area for a specific broadband project). Which basis is most appropriate depends on the question being asked.

The initial focus of this study was to collect metrics at the community/town level for the sample of 14 communities, but which could be extended to the 351 communities in Massachusetts. Focusing on town boundaries may help address particular constituencies and exploits natural boundaries for the collection of some sorts of basic infrastructure data, but may poorly correspond to broadband service market boundaries or the boundaries of broadband projects.

Ultimately, what matters to end-users is whether they can get broadband where they live, not that broadband may be available to some of the residents in their town. When it comes to broadband services, availability can differ substantially over the footprint of a town. For DSL, the further form the Central Office, the lower the data rate you may be able to obtain and, if sufficiently far

[33] We focus on facilities-based competitors because this provides insight into infrastructure availability. While retail-level competition is important, it does not provide an adequate basis for competition in the absence of regulatory constraints in the absence of actual or credible facilities-based competition.

away, you may not be able to get DSL service at all. (In contrast, the quality of POTS services do not typically vary over the service area.) For FTTH, the variability may be even greater: services may be available on one block but not an adjacent one. And, for wireless services, even nominal availability of 3G services in an area does not mean that users within that area can reliably get 3G service. Thus, it is preferable that data focus as much on point-availability of service as possible, rather than on some larger (and inherently arbitrary) geographic area definition.

The above notwithstanding, it often will be necessary for practical (including cost) reasons to collect data based on area aggregates. A number of candidate options are available, including neighborhoods, zip codes, and Census tracts. Of these, we concluded that Census tracts offered the best match of availability with third party data (e.g., to stratify data by race, income, or educational attainment) and yet reasonably small size. Identifying neighborhoods consistently across communities and over time proved intractable, and zip codes are often too large. For the 2000 Census, there were almost 1,400 Census tracts in Massachusetts. Census tracts are designed to include approximately 4,000 people (although some have as few as 1,500 or as many as 8,000.) Thus they strike a good balance between granularity and manageability. However, there are a couple considerations to keep in mind when using Census tracts. They sometimes spread across town boundaries, which would make it more difficult to aggregate the data to the town level. Also, although they are intended to be stable enough for measuring long-term trends, it is still possible for them to change over time.

## 2.4. Challenge of composting multidimensional metrics

From the above, it is clear that there are multiple dimensions along which it might be reasonable to define access availability: with respect to where access is available (public terminals or at home, by town or some other area aggregate), in terms of the quality of services available (tiered by peak data rates), in terms of price/affordability, or in terms of the range of competitive choices. Along a single dimension it might be relatively easy to rank communities in terms of which are better served.

For example, at the lowest level of home access, one may ask whether a community has *1B or better* broadband service available. Communities where there is no 1B service available anywhere would unambiguously be considered unserved. While communities wherein *1B or better* service is generally available to every home in the community would unambiguously be considered as served. The share of homes that could be served by broadband (i.e., are passed by broadband services, even if the households may not subscribe) offers one obvious way to rank order communities in terms of their degree of broadband service. With such an approach it might be possible to identify communities as unserved by broadband if only a small share of homes have 1B or better service available from a single provider, where the share is chosen so as to eliminate examples of incidental availability of broadband (e.g., less than 10% of the homes).[34] Likewise, it is reasonable to conclude that communities with better than 1B service options

[34] Conversely, it does not seem reasonable to set the threshold too low (say, for example, less than 1%) because there are likely to be a small fraction of households in some communities that are extremely expensive to serve and verifying that very few such exceptions to universal service availability exist is likely to be expensive. Thus, it ought to be reasonable to claim effective *universal availability* with something less than 100% of homes served.

available to at least some homes are better served than communities with only 1B service to the same number of aggregate homes covered.

Another dimension to consider is the quality of broadband service available: communities where better than 1B service is generally available are better served than communities with only 1B service. Similarly, it is possible to consider the range of choices available to consumers from facilities-based providers.

The composition of rankings becomes more difficult when we combine rankings or metrics that have different implications. Consider the following examples and ask whether you are convinced which community is better served (and ought to have a higher Broadband Community Index):

SAMPLE HYPOTHETICAL COMMUNITIES

A: FTTH available to 80% population, no broadband to 30% of households

B: 1B available to 80% households and 2B public access within 1 mile of everyone

C: 1B available to 95% households, no broadband remaining 5%, no 2B anywhere.

How would you rank these communities? One plausible ranking might be C>B>A if the focus were on home access to broadband, but A>B>C if the focus were on availability of next generation broadband services. One could further complicate the examples by considering competition and affordability options. However, these should suffice to demonstrate the point that the appropriate ranking depends on the context of the question being asked.

## 2.5. Identify criteria for selecting metrics

The choice of metrics needs to satisfy four criteria:

1. Informative: tell you what you need to know (depends on context)
2. Simple: intuitive, easy to implement/explain/use
3. Feasibility: implementable (strategy for measurement exists)
4. Cost effective

Each of these is discussed further in the following sub-sections.

### 2.5.1. Informative

The metrics need to be informative which means that they provide a discriminating way to assess the level of broadband across communities or other areas. That means that the metrics need to be applicable both to poorly-served *and* well-served communities.

The focus here is on broadband availability, although other concepts like adoption or availability of complementary goods (e.g., home computer ownership) may be necessary to adequately interpret the data. For example, if your question is to better understand why broadband is not being used as intensively in one community as another, availability is only one factor that impacts that outcome. Thus, what constitutes being informative will depend on the context of the question being asked.

To be informative, the metrics must be well-defined and fully specified. Users should not have to guess what the units are or the basis for computation.

Finally, informative metrics are analytically robust so they ought to make sense when aggregated, resulting in logical and consistent rankings. For example, if you have three communities with service rankings S(A)>S(B)>S(C), you would not expect to see S(B+C)>S(A).

### 2.5.2. Simple

To be useful, the metrics need to be understandable for those who will need to interpret them. This helps if the metrics are intuitive. If folks cannot understand the point of the metrics quickly, then that may mean they need to be simplified.

One way to simplify the metrics is to keep the number relatively small. In assessing broadband, more than one measurement or metric is needed, but the total number of measurements that might be made for a community probably should not exceed 30 and 5-10 may be a more reasonable goal. Limiting the number of metrics has several practical benefits. It lowers data collection costs and eases interpretation, but does limit the range of questions that may be answered.

### 2.5.3. Feasibility

Useful metrics must be feasible to implement. That means that the data is available. The measurement strategy is well-defined and where proxy metrics need to be used, there is a reasonable expectation of accuracy. For example, in the early days of broadband, when adoption was availability-limited, availability could be used as a reasonable proxy to infer adoption rates. Over time, as availability has expanded, variability in adoption rates has grown and availability offers a much less reliable proxy with which to infer broadband penetration.

The likely measurement error needs to be considered when evaluating alternative metrics.

### 2.5.4. Cost effective

Finally, and closely related to the goals of simplicity and feasibility, it is important to conserve resources. This includes minimizing the direct costs engaged in the measurement exercise, as well as the indirect and spillover costs. For example, the time it takes to complete surveys is an important consideration. The longer the time, the greater the opportunity cost for participation and the less likely folks will be to complete the survey. More difficult to estimate are potential spillover costs if the measurement strategy induces inappropriate strategic behavior (e.g., attempts by folks to game the measurement effort or squabble over its interpretation). We discuss such adverse implications further below when we consider the applicability of a Broadband Community Index.

What constitutes an appropriate level of costs depends on the accuracy required. Typically, spending more resources will allow better measurement (although not always).

# 3. Metrics

We first discuss different types of metrics, and then recommendations for appropriate broadband metrics for measuring availability in Massachusetts.

## 3.1. What types of metrics

There are many types of metrics that could be used to assess broadband availability. These include both absolute and relative metrics, and metrics that have different time horizons or analytic purposes.

### 3.1.1. Absolute metrics

The most obvious metrics are absolute metrics that identify what the level of availability of service is. This includes measurements that indicate the number of households that are not served by any broadband provider or the counts of broadband lines in service for different types of technologies. The absolute availability metrics also need to consider the range of choices and the price/costs associated with selecting different choices. For example, what is the least cost option available and what is the option most commonly selected?

### 3.1.2. Relative metrics

To ease interpretation, relative metrics, which are normalized, may be more informative. For example, the share of houses that are unserved normalizes for the size of communities. Such normalizations are important when trying to make comparisons across communities. It is also important to look at the within-community disparity levels. For example, what is the variation between the best served and worst served parts of a community?

### 3.1.3. Needs assessment

It may also be relevant to conduct a needs assessment, by adjusting metrics to account for per capita income, "eReadiness" ranking,[35] or likelihood of adoption. For example, if one had two communities that are equally poorly served but one that was more likely to adopt sooner (e.g., higher incidence of home computer ownership), then the availability-need for the early-adopter community might be higher.

### 3.1.4. Cost to address

It may also be worthwhile considering the relative disparity from the availability goal and the costs of achieving that goal. Assessing such costs may involve considering differences in

[35] Multiple frameworks have been defined for assessing the quality of the Internet environment across markets and communities. For example, the Economist Intelligence Unit (EIU) prepares eReadiness rankings by collecting a number of measurements for 70 countries that measure such things as the "connectivity environment," or "underlying social and cultural attitudes surrounding Internet adoption." The Economist defines "e-Readiness [as] a measure of the quality of a country's information and communications technology (ICT) infrastructure and the ability of its consumers, businesses, and governments to use ICT to their benefit." See "E-readiness rankings 2008: Maintaining Momentum," a white paper of the Economist Intelligence Unit, prepared in cooperation with the IBM Institute for Business Value, April 2008, available from: http://a330.g.akamai.net/7/330/25828/20080331202303/graphics.eiu.com/upload/ibm_ereadiness_2008.pdf

infrastructure costs (e.g., housing density, lower density usually means higher cost) or availability of complementary infrastructure (e.g., proximity to long-haul fiber).

#### 3.1.5. Trends

Finally, to put availability measurement into perspective, it is necessary to track metrics over time. Some metrics may relate to progress (how fast has gap narrowed?) or trends (is availability improving?). This implies thinking about the costs and feasibility of measurement repeatability so a time series of informative data may be collected.

### *3.2. Defining broadband availability metrics*

Perhaps the single most important metric is the percentage of households in a community (however defined) for which broadband service is available.

This is the main metric that the E-NC Authority in North Carolina uses to measure availability. They track this value for all of the counties in the state. Their most recent annual report (2006-07)[36] shows that many counties have availability rates in the range of 70-89.9%, but 21 counties have access below 70%, and 4 counties have access below 50%. Those latter 4 counties are the counties targeted in their most recent RFP. As stated in the RFP:[37]

> "Access to high-speed Internet service for all households in North Carolina is a primary goal of the e-NC Authority. The incentives strategy of the e-NC Authority is based on a tiered approach, focusing on counties with less than 50% access to households first, then those with less than 60% access, then less than 70% access. The e-NC Authority believes it is critical for all counties to achieve at least 70% access as soon as possible. There are currently 21 counties with less than 70% access (per e-NC's December 2006 data). However, with the funding available, this Request for Proposals covers just the counties where less than 50% of their households have access to high-speed connectivity."

One may be interested in the density of public access points, which provide an alternative means of Internet access for people who cannot receive Internet access at home. In addition, having a choice of broadband providers is also desirable. The following table shows some additional metrics of interest:

| | |
|---|---|
| Access with travel | What % of households are within X miles of a public access point?<br><br>Density of public access points (per capita or sq. mi.) |
| Basic access | For what % of households is standard broadband (2B) available? |
| Market competitiveness (1) | What % of households can get broadband from more than one |

[36] http://e-nc.org/pdf/e-NC_Biennial_Report_06-07.pdf

[37] http://e-nc.org/pdf/Connectivity_RFP_9-14.pdf "E-NC Authority Request for Proposals (RFP) Connectivity Incentives Grants, Supply Side", Sep. 14, 2007, p. 3.

| | |
|---|---|
| | provider? |
| Market competitiveness (2) | What % of households can get broadband from more than two providers? |
| Next-generation broadband | What % of households can get next-generation broadband (3B)? |
| Competitiveness of next-generation broadband | What % of households can get next-generation broadband from more than one provider? |
| Equity | What is the lowest % availability found in an individual neighborhood? (The purpose of this metric is to identify where there are individual neighborhoods which have significantly lower levels of availability than the rest of the community.)<br><br>Are there any socioeconomic classes (including racial and ethnic groups) which experience lower levels of availability? |
| Access for teachers | What % of schools have Internet connections for teachers and administrators? |
| Access for students | What % of schools have Internet connections for students? |
| Access for businesses | Is next-generation broadband available to businesses? |
| Competition in business market | Is next-generation broadband available from more than two providers? |
| Cost of access for business | What is the monthly cost of next-generation broadband for business? (lowest available anywhere, and most common) |
| Access for hospitals | Do all hospitals have access to next-generation broadband? |
| Next-generation access for schools and universities | Do all schools and universities have access to next-generation broadband? |
| E-government | Does town government provide services electronically? |

For example, in Massachusetts the public libraries already provide public access to broadband in most communities across the state. As of June 2006:

- Almost all public libraries in MA (359 out of 370) had Internet connections and terminals for public access[38]
- 58% of libraries had a T1 connection, 39% had a cable modem connection, and, 11% had DSL.
- 50% of libraries were providing wireless Internet access.
- In addition, as of June 2007, about 83% (291 out of 351) of the town halls in Massachusetts had broadband connections, according to the John Adams Innovation Institute[39].

Once you know the status of service availability, it will be interesting to measure other things such as:

- Adoption rates: how many households are currently subscribing to broadband?
- What are households paying for broadband?
- How much and what broadband services (applications) are consumers using? Are consumers teleworking? Are consumers accessing eGovernment services? Are they engaging in ecommerce?
- How is broadband being used in the schools? Is it impacted by home use?
- What broadband services are available in the community (e.g., do local businesses provide services on-line? Are there adult education and community user-groups on-line?)
- How aware are consumers of the options available? Do they know which providers provide broadband service and what is their level of Internet sophistication?
- What do consumers think about the quality of broadband service available? How best could it be improved (price, speed, options)?

The range of potentially interesting questions is endless, and it is common that the more you know, the more you may wish to know.

The above questions focus on how broadband is being used, which is relevant if you wish to understand the economic impact of broadband and how best to improve the nature of broadband services in the future. After all, we are only interested in broadband availability to the extent that we think these are services are necessary and will be widely adopted and used. Realization of the benefits of broadband depend on it being used, and widespread adoption results in externality and scale benefits. Thus, the larger the market for broadband subscribers, the greater the

[38] Commonwealth of Massachusetts Board of Library Commissioners, "Massachusetts Public Library Data, Electronic Services FY2006 Data (July 1, 2005 – June 30, 2006)", distributed June 2007, accessed at http://mblc.state.ma.us/advisory/statistics/public/repelec/elec06_report.pdf.

[39] John Adams Innovation Institute, research and analysis at http://masstech.org/broadband/research.html

incentive for content and application providers to offer valuable content and the greater the market over which to recover fixed costs (and hence, the lower are average costs).

Observing how broadband is being used will require more than data on the location of infrastructure. While service providers do have detailed insight into their own subscribers, they may have only a limited idea what subscribers of other providers are doing and justly regard their subscriber data as confidential. While providers may be willing to provide aggregate data on subscribership and usage patterns, consumer surveys may be needed to collect more detailed data on subscriber patterns.

### *3.3. Some sample metrics*

In the *Broadband Metrics Best Practices Report*, we suggested a range of possible broadband availability goals:

| Goal | Title | Description |
|---|---|---|
| | | ***Achieving Ubiquitous Availability*** |
| L0 | Town government is on-line | • *1B* broadband (say a T1 line) is available to at least one building (say Town Hall) in all 351 towns in MA. |
| L1 | Public access BB is available | • *1B* broadband is available in every public library and public school, with no household more than 5 miles from a public-access terminal |
| L2 | Ubiquitous BB available | • *1B* broadband available to (almost) every HH (95% availability in every town) from at least 1 provider |
| | | ***Keeping BB on track*** |
| L3 | BB adoption on track | • BB adoption rates are on par with national average.<br>• $/Mbps/month for average, best, and entry service on par with national averages<br>• Within state differences on par with peer states. |
| L4 | BB is best in class | • BB availability and adoption rates for higher quality BB services (*2B, 3B, 4B*) are on par with national averages. |

- Within state differences on par with peer states

These goals provide one way to think about the hierarchy of broadband standards that may be considered in assessing the status of broadband infrastructure. At the lowest level, there is the goal to get town government on-line. Achieving that goal at least provides a point of contact within each community for email and Web access to enable data collection. The L1 level is intended to signify that broadband access is generally available to the public, even if not at their residences and even if they do not have a computer. Finally, and generally the focus of most goals to expand broadband availability is L2 – at least first generation (1B) broadband available ubiquitously, to every household in the community. While suggesting a natural hierarchy of progress, these goals are independent. For example, it may be the case that ubiquitous household coverage is achieved before adequate public access is available. However, we believe that both are needed – as noted above – to ensure that broadband is, indeed, ubiquitously available.

Achieving availability of ubiquitous broadband is only the first step. For broadband to deliver economic benefits, it has to be adopted and used. The second set of goals focus on keeping broadband development on track and on par with other peer states. L3 is designed to provide a basis for benchmarking progress in Massachusetts against other states. Obvious metrics to consider are such things as the statewide average adoption rates and $/Mbps/month for the typical subscription provided. It is also important to look at the level of disparity across the state and to consider this in comparison with others. Looking at broadband adoption rates by income class, rural/urban, or across counties may provide useful indicators for this.[40] One might also look at such things as the level of customer complaints per broadband lines or the average peak data rate (proxy measures of service quality). Finally, L4 is intended to compare the progress of broadband for best-in-class services. While it may take a long time, if ever, for every home to have 3B/4B "FTTH" quality infrastructure and broadband services available, some areas already have it. L4 is intended to focus on ensuring that progress in Massachusetts for those early-adopter communities is on par with other leading states. Possible metrics for tracking L4 progress might include such things as the share of households with FTTH service or services with peak data rates in excess of 50Mbps, or investment-per-capita in next generation networking infrastructure.

With each level, multiple tests/measurements will be useful to more fully characterize the states achievement. The following table suggests some useful measurements for each level:

---

[40] For example, you could compute the sample variance in broadband adoption across communities or counties in the state, or $S = \frac{1}{N-1} \sum_{i=1..N} \left(s_i - s^*\right)^2$ where $s_i$ is the adoption rate in each of the i=1..N communities/counties, and s* is the average adoption rate.

| | Level of Service | Description of Measure | How to Measure |
|---|---|---|---|
| L0 | No BB available in town | No service provider, no BB available to government, no public access BB terminals in town, No businesses with BB | |
| | | -- No provider selling BB data service to enterprise or mass market customers | Number providers selling BB in the community (identify, FCC data) |
| | | -- No businesses/enterprises with data service greater than 500Kbps (downstream). Using data service | ?? |
| | | -- No government office, school, library, hospital, police/fire with BB service | Identify by category (survey critical infrastructure?) |
| | | -- No publicly-accessible BB terminals (library, school, or coffee shop) | Libraries with BB, Schools with BB, Public coffee shops (Starbucks, other?). Counts and GIS-coded by address. |
| | | -- No service in adjacent community (lower the score for adjacent, lower score for community) | [Use other measures. All adjacent communities.] |
| | | -- Every HH is within x miles of publicly-accessible BB terminal (in library, coffee shop, town office, school, etc.) | Density of public access points; Share of HH that more than x% |
| | | | |
| L1 | BB public access | BB available (nearly) ubiquitously in public-accessible buildings | |
| | | -- Share libraries with BB | Number, % of total, Public access terminals/Capita, Connections (Sites)/Capita, Density of connections |
| | | -- Share town offices with BB | Compare with national average, state average, community average |
| | | -- Share schools with BB | Longest distance travel |
| | | | |
| L2 | BB residential available | BB available (nearly) ubiquitously to all HH in community | |
| | | -- Number of providers in each zip code | |
| | | -- Number of homes without even one provider | |
| | | -- Share of homes covered by technology (DSL, cable, FTTH, wireless) | |
| | | | |
| | | | |
| L3 | | BB benchmarking against other states | |
| | | -- BB availability on average (take-rates) are at or above national average | Average penetration. |
| | | -- Competition | |
| | | -- QoS good and improving | Average data rate. Coverage-weighted share served by highest rate. |
| | | | |
| | | | |
| L4 | | BB infrastructure health and high-quality | |
| | | -- FTTH penetration is high | FTTH deployment matches or exceeds comparable states |
| | | -- Essential businesses have capacity they need | BB capacity in state (traffic metrics) are higher than average |

Each of the qualitative level goals identified above could be quantified by setting specific thresholds. For example, "ubiquitous household" coverage implies that there is no community within the state with more than 5 percent of its households unable to obtain broadband. Proof that this is the case could be by sampling, where the size of the sample would depend on the size of the community.

A set of related metrics that could be used to measure increasing levels of broadband service that could be used to produce a color coded map (Red to Pink to Yellow to Light Green to Green implying better broadband service) are suggested in the following table:

| | |
|---|---|
| 0 | No BB in town (BB wholly unavailable so extreme boundary condition. |
| 1 | Some % of folks can leave home and travel moderate distance to access BB (this is better than no BB at all, but may still leave a sizable share of population with |

| | |
|---|---|
| | no service (distance required to travel to public access is excessive) and inadequate service for the rest if there is no residential home access). May have threshold % -- if more than x% must travel more than y miles for public access than community is considered "underserved." [Code Red] |
| 2 | Small number must travel far to access BB (while still no res BB, most folks can get if they want from public access). [Code Pink] |
| 3 | Share of homes with BB service available at home is large (most folks have 1B available from at least single provider) (Rest may or may not have other options and how to compute is an issue -- e.g., travel time to nearest public access point for those without BB at home, or average travel time over all residents). [Code Yellow/Pink] |
| 4 | Share of homes with BB service unavailable at home is small (substantially achieved "ubiquitous" BB availability from at least one provider) [Code Yellow/Lt. Green] |
| 5 | Share of homes with BB service from 2 or more providers is large (most consumers have choice but significant number still do not) [Code Lt green] |
| 6 | Share of homes with BB service from only 1 provider is small. (Only a small share of homes lack choice for BB service) [Code green) |
| 7 | Share of homes with FTTH is moderate (availability of advanced services for moderate group) [Code darker green) |

The above demonstrate the importance of considering how conditions vary over the population in an area. Thus, measuring averages and marginal conditions are important. Goals/targets might be made operational by specifying both an average target and worst case constraints.[41] Examples of such measurement standards include:

[41] Let s be the target measure (higher is better). Then average measure in community, s*>s^ and min(si)>s0 where s0 is a minimum threshold, such that s^>>s0. This establishes the dual commitment to achieve a minimum level of service on average, while also committing to a minimal floor of service availability for everyone.

| | Metric/measurement | Average | Worst (best) case |
|---|---|---|---|
| 1 | Distance travel to public access point (pop weighted avg miles) | Avg. distance<y1 | Max distance<z1 for not more than w1% pop |
| 2 | Density of public access points (per capita-sq mile) | Density>y2 | |
| 3 | HH served by xB | %HH (or Pop) served>y3 | %HH unserved<z3 |
| 4 | HH served by 2B+ (e.g., Share FTTH) | %HH (or Pop) served>y4 | %HH unserved<z4 |
| 5 | Affordability (Average price paid, line weighted) | $/month<y5 | $/month for best offer<z5 |
| 6 | Quality (line weighted average peak data rate) | Average Mbps>y6 | Mbps>z6 for not more than w6% of pop |
| 7 | Choice | %HH with 2+>y7 | %HH with only 1<z7 |
| 8 | Essential facilities | %hospitals, police, government offices w/ xB>y8% | %hospitals, police, government offices w/0 xB<z8% |

### 3.4. Proposed survey methodology

As noted in the introduction, one way to collect the data for the metrics noted above is by conducting end-user surveys. The design of the survey (what questions to ask, the language and organization of the survey, who to ask, how to implement) have important implications for the quality and cost of the survey.

As already noted, if the principal goal is to assess availability, then the best approach is a GIS-mapping strategy targeted at service providers. End-user surveys may offer a valuable tool with which to complement the GIS-mapping survey. To allow statistically robust inferences, implementation of end-user surveys will require relatively large samples, are complex to design, and are not inexpensive (although less expensive on a per-survey basis than developing a detailed GIS-map). However, conducting statistically valid end-user surveys will be important to learn about adoption/usage behavior, consumer expenditures/prices, and consumer attititudes and desires. Thus, such surveys should be part of a best-practices broadband metrics approach.

In the near term, smaller and lower cost surveys may be used as a quick and dirty way to develop data on broadband availability and other factors, however, the reliability of such measurements will be suspect.

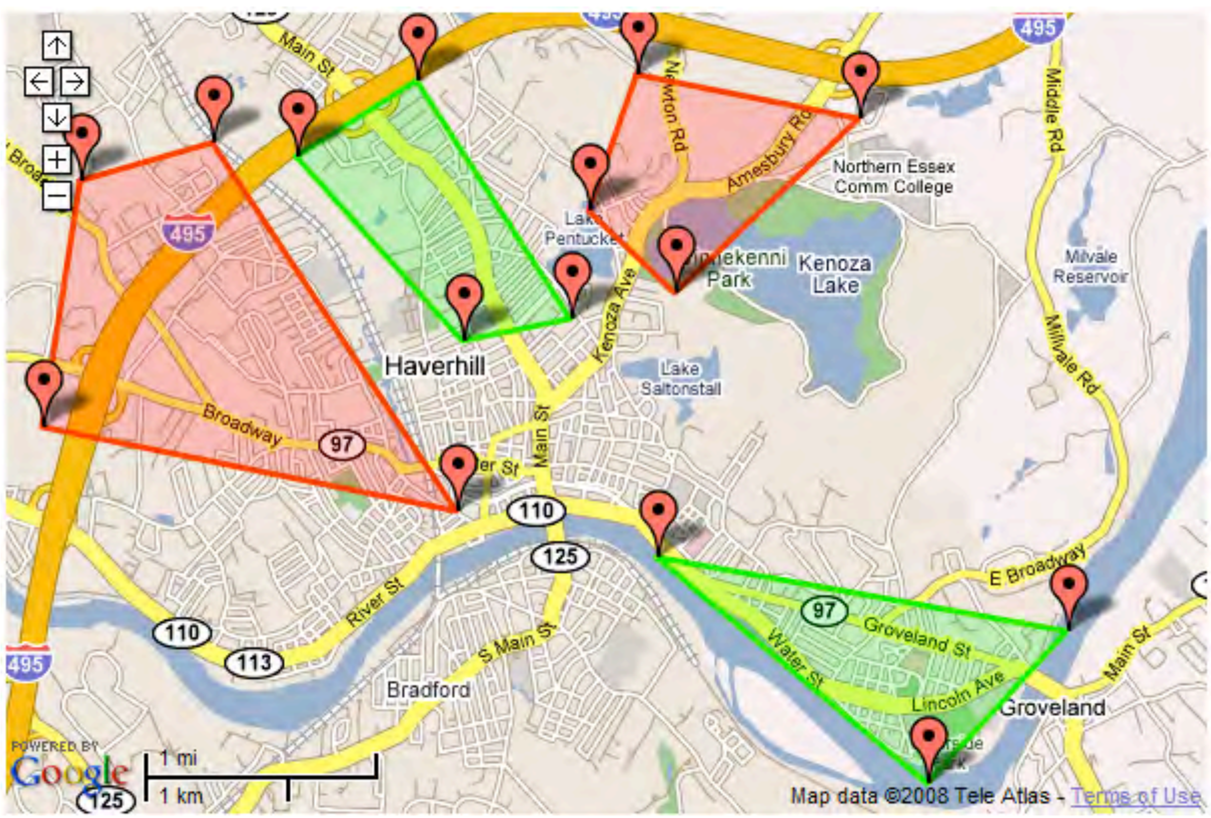
The typical end-user survey approach focuses on targeting a representative sample of consumers (by demographics, location, etc.). Within the scope of this project, we had inadequate time or resources to implement a statistically valid consumer survey approach. While we believe such surveys ought to play an important role in future broadband metrics programs, especially because these offer one of the best ways to observe valuable information about end-user perceptions and expenditures.

An alternate approach to the typical end-user survey is to find a sub-sample of broadband-knowledgeable folks who are willing to serve as volunteer broadband researchers and complete a more detailed survey. For example, one could ask the town IT manager or a local broadband advocate to tell you what services are available in the town and where they are available. It is reasonable to believe that this latter strategy might be able to make use of a more extensive survey than one could reasonably expect to have completed by a random sample of end-users. Also, such a focused "broadband expert" survey could be much smaller, would provide useful guidance for the design of subsequent random consumer surveys, and would help provide an initial core for community-based organization of future metrics collection efforts.

In light of the advantages offered by such a "broadband expert" survey, we designed a detailed survey to gather all of the information which would be needed to estimate the desired broadband metrics. The survey is included in an appendix to this report. It requires knowledge of broadband availability at a detailed, neighborhood level.

We also considered developing a visual tool, based on Google Maps, which allows users to enter this information on a map. The attached figure shows what such a tool could look like. It allows a knowledgeable user to mark different regions of a town and color-code them according to their broadband status: red indicating where broadband is not available, and green indicating where it is available. This information would be saved in a format suitable for manipulation with a GIS system, and could also be combined with demographic data. For example, the GIS could use US Census population data to estimate what portion of households lie within serviced and unserviced areas.

**Figure 1. Example of a Google Maps-based tool for collecting information on broadband availability**




According to the original project plan, it was expected that we would submit the survey to a panel of IT-savvy folk across Massachusetts in March 2008. In light of what was going on at the time, doubts that the appropriate individuals with the right expertise to answer the questions could be identified, and concerns about the burden imposed on any such individuals who might be identified, it was decided to not go ahead with the survey.

While such an approach may prove useful in other contexts, it is unlikely to be an appropriate substitute for statistically valid (large sample) end-user surveys.

## 3.5. Applicability of a Broadband Community Index

As noted in the introduction, given a collection of measurement data for each community and a set of defined metrics, it would be possible to compute estimates of the metrics. This collection of estimates could then be reported or used to make a variety of decisions, or could be combined to construct a weighted index.

While theoretically plausible, there are a number of practical concerns as to the validity or appropriate use of a Broadband Community Index. Its chief benefit is that it provides a summary statistic for compositing the impact of multiple dimensions. This is also its chief drawback as already noted in the discussion regarding the context-dependence of ranking multi-dimensional metrics. This suggests that if such an index is constructed, it ought to be constructed with a clear focus on the question to be addressed. The weights and composition of metrics that comprise the index may change as the analysis context changes. For example, if the focus is on broadband availability irrespective of type, then it may not be relevant to distinguish among different types of broadband; however, if the focus is on the availability of more advanced services then metrics indicating the availability of FTTH, ADSL or more advanced cable modem services should be included in the index.

Even assuming an appropriate index can be defined, it is doubtful that such an index would be appropriate to use as the principal basis for allocating state broadband development funds. There are several reasons for this. First, if applicants know that the index will be used to allocate funds, then that will create a strategic incentive to seek to manipulate the measurement results, increasing the measurement challenges.[42] Second, if the goal of the public broadband funding is to produce the largest welfare impact per dollar invested or to maximize the expansion in the number of households covered, it is doubtful that the index will provide a good way to rank potential projects. For example, it is quite possible that a project with a higher impact in terms of expanding the number of consumers served by broadband (on a cost-adjusted basis) may be targeted at a community which is better served, as measured by the index. The committee making allocation decisions may find reliance on an index and simple quantitative cut-offs as overly restrictive and unnecessarily contentious. A more appropriate and common way to select among multiple projects is to announce funding program goals and application criteria, and then allow the committee to evaluate these with some degree of flexibility to compare projects considering the range of project submissions received, the portfolio implications of funding multiple projects, and the opportunity to seek additional clarifying information. Certainly, the data to compute an index may be appropriate to consider in the project application evaluation process, and also in subsequent evaluation of the impacts from funded projects. However, we

[42] The incentive to manipulate data may be addressed in part by computing the index as of some historical date.

recommend down-playing the importance of a quantitative ranking as per any index in the decision process for allocating broadband development funds.

While creation of an index may be appropriate in the future and in a specific context, we do not believe the BCI-approach represents the best path forward for assessing the health and status of broadband services and infrastructure in the state. Reliance on a BCI would be contentious as it would induce disagreements over the appropriate weighting of component metrics and would suggest a level of measurement precision that is unlikely to reflect relevant data constraints. Some of the elements that it would be desirable to include may be more qualitative (subjective) or categorical, and so not readily translatable into a scalar component.

The problem posed by a BCI is akin to the problem of trying to evaluate college applicants solely on the basis of a standardized test such as the SAT or to assign high school students a 0 to 100 numerical "score." Instead colleges like to evaluate an application package that includes a diverse array of information, including SAT scores, high school transcripts, letters of recommendation, and student essays.

The BCI concept proved valuable as a conceptual tool for motivating aspects of the broadband measurement challenge. It helps force attention on the fact that broadband service availability reflects a continuum and requires consideration of multiple aspects/attributes (e.g., it is not appropriate to just focus on a single scalar metric such as the percentage of households passed or penetration rates). Further consideration also suggests the inherent arbitrariness of weights and the difficulties of aggregation (defining appropriate areas for measurement). These problems are innate to the metrics challenge and cannot be eliminated.

In light of the limitations of the BCI concept, we recommend that a report card approach be adopted to measure broadband services. We discuss this approach further in the following chapter.

# 4. Toward a broadband report card

In this chapter, we offer an example of a broadband report card that could be compiled for a community. While we believe this example could be adopted as a practical solution as is, our intent in presenting it is less prescriptive than illustrative. We believe it represents a useful starting point that may be modified on the basis of on-going discussions among those actually tasked with implementing the broadband metrics project. As such, we believe it important to explain our rationale in establishing the various design choices.

We present our proposed report card in the following two sub-sections. First, we identify the core elements/features that motivated its design. Then, we present the actual metrics and discuss ways in which it may be motivated or expanded in the future.

## 4.1. Key report card features

We envision a broadband metrics report card that could be compiled for each community in Massachusetts as being analogous to a student's report card. Our proposed broadband report card embodies the following features:

- Few metrics: 5-10 grades

  Limiting the number of reported metrics offers a number of important advantages. It eases comprehension, lowers costs, and reduces the points of debate. Having too many measures risks losing focus on what is important.

- Coarse scoring: red/yellow/green

  We recommend using a coarse scoring to reflect the inherent data uncertainty and to reduce the points of debate. The goal is to provide metrics that are clear and relatively unambiguous.

  We interpret red to mean "unacceptable", yellow to mean "potential issue", and green to mean "acceptable."

  Color coding provides a quick way to visually summarize a complicated mix of data quickly. The sorts of comparative judgments we would expect to see are "where are the red spots?" or "is the map becoming more green over time?" or "where are the yellow areas where we need investigate further or may have an issue?"

- Flexibility

  Our proposal is illustrative, not prescriptive. While we do propose a specific set of metrics or "grades" to be included in our report card, we expect that grades may be added or removed as needed and as better data becomes available. Just as it is feasible to compare student report cards with more or less course grades to identify better and worse performing students, so the broadband report cards may evolve over time.

  Some of the grades (metrics) that comprise the report card may be BCI-like indices or numeric measurements (e.g., % of homes passed by broadband). Other grades may be more qualitative or subjective.

  Additionally, the report card framework with multiple grades allows users to apply a variety of implicit weighting filters when sorting or analyzing the data. Some contexts and some users may focus on a single metric/grade, while in other situations, an overall evaluation may be more appropriate.

- Readily comprehensible

  The purpose and use of a grade ought to be clearly articulated and intuitively easy to understand.

- Simple Answers

  The grading system we propose is based for the most part on a series of Yes/No questions (with third option being, Unknown), but scalar data may also be included if it is available. Keeping the answers simple helps reduce ambiguity and, hopefully, reduces the challenges of data collection.

Our goal is to allow scoring for the report card to be mechanical and judgment free to reduce incentives for strategic behavior or concerns about fairness.

## 4.2. Broadband report card

In developing our proposal, we build on the general framework for establishing goals for broadband assessment first presented in the *Broadband Metrics Best Practices Report* and discussed again at the beginning of section 3.3 above. This framework described a set of 5 goals which we mirror in a set of 5 course grades, which define a collection of key measurement concerns. The first three focus on ensuring universal availability of broadband services and are the most relevant given today's concerns and the current status of broadband market development. The last two focus on benchmarking the health of broadband services over time to minimize inequities within the State and to ensure that Massachusetts is on-track to be best-in-class. Because the last two "grades" are more future-oriented and depend, in part, on having access to the data required to score the first three "grades," these last metrics are not completely specified.

### 4.2.1. "L0": Local government is on-line

The first grade is intended to measure the availability/use of broadband by local government for its internal use and communications. This is important because it provides insight into the progress being made toward implementing eGovernment.

The connection to ensuring mass market broadband availability here is indirect, but we believe that as a minimum requirement, town governments ought to be on-line. Ensuring this is the case provides a point of contact for eGrass-roots organizing.

Over time, as broadband availability goals are achieved, we expect local governments to expand their eGovernment activities and use the broadband Internet to communicate with citizens and run government functons more efficiently. The questions may be expanded to track such developments in the future.

Our initial proposal for scoring this grade is based on the following three questions:

- Q01: Does the town hall have BB available for its internal use?
- Q02: Is there a municipal intranet linking (most) government offices, libraries, schools, hospitals, police/fire?
- Q03: Does the town maintain an active municipal website?

Scoring for this grade should be as follows:

- Green: Answer to all three questions is Yes.
- Yellow: One No or Unknown answer
- Red: Two No answers

### 4.2.2. "L1": Public Access to BB

This second grade measures the status of public BB access options. As noted, ensuring adequate access to public broadband terminals ensures that broadband is available even to families without computers at home, or who may need some help in using broadband, or who either cannot get broadband at their homes or choose not to subscribe.

Our initial proposal for scoring this grade is based on the following three questions:

- Q11: There are 1 or more public access terminals with BB access in the community?
- Q12: Public schools have BB access for educational (and internal) use?
- Q13: Public libraries have public access terminals with BB access?

Scoring this grade should be as follows:

- Green: Yes to all three questions.
- Yellow: Unknown to one or more questions.
- Red: No to one or more questions.

As noted earlier, this metric may be expanded by considering such things as the density of public access terminals (Terminals/Pop-Area) or by measuring the average or maximum distance that someone might need to travel to get to a public access terminal. It is also possible to consider quality (what broadband service supports the public access terminals? What applications are supported?) or price (what are the libraries paying for access?). While such additional data would be interesting to have, we did not include it in our proposal because of data constraints and in the interests of simplicity.

### 4.2.1. "L2": Ubiquitous BB Availability

This is the single most important grade/score for assessing mass-market broadband availability. We believe each of the proposed questions here ought to be separately scored and reported, and then an overall score ought to be reported.

Each of the questions focuses on a different aspect of BB availability, and these are arranged hierarchically in order of importance. The specific cut-off points included are provided for illustrative purposes and other levels may be chosen (e.g., national averages).

The following questions should be included in this grade:

- Q21: Are 1B broadband or better ("1B+") services available to 95% or more of the homes? (Basic availability)
- Q22: Are 1B+ available from more than one facilities-based provider to 60% or more of the homes? (Competition and consumer choice)
- Q23: Are 2B+ available to 60% of homes? (Quality of service). This metric may be expanded to include other measures of service quality such as customer complaints or outages.
- Q24: Is price for 1B service at or below state average?

Scoring of this should be as follows:

- Green: Q21 is Yes, at least one more Yes for Q22 or Q23, and zero No answers. Logic here is that Q21 is the most important and trumps the other questions. While unknown answers are acceptable for other questions, a single No moves the score to a Yellow.
- Yellow: Q21 is Yes and at most one No; or Q21 is unknown and zero No answers.
- Red: Q21 is No; or Q21 is Yes and two or more No answers; or Q21 is unknown and at least one No answer.

The logic for scoring in this way is to emphasize the goal of universal coverage. We are skeptical that a higher standard than 95% coverage would be warranted, but it might make sense to lower this standard to say 90%. This would expand the range of communities that might be coded "Green." We do not have a strong feeling about whether Q22 (competition/customer choice) or Q23 (quality, which here is measured by peak data rate) is relatively more important. We have nominally put Q22 ahead of Q23, but anticipate that these would often be complementary since the markets with the highest quality services are also likely to be the markets with the most competition. The choice of 60% as a threshold is just to suggest that "most of the customers have a choice" but to allow for the possibility that the footprint of competing facilities-based providers is non-overlapping. The focus on facilities-based providers is to control for the fact that there are many markets with DSL resellers. While such competition does offer valuable consumer choice, it is a different sort of choice than exists when there are multiple facilities-providers. The latter type of competition is more important for assessing the extent of infrastructure competition and diversity.

We put Q24, the price/affordability metric last because this data is likely to show the least variability across communities and is often the most difficult to get without consumer survey data.

We would anticipate that the data rate thresholds and the share of households with multiple facilities-based choices would increase over time and that the adjustment in these thresholds may me marked to changes in national or statewide averages.

Over time, we would expect that scalar data on the level of coverage, the average data rate, the average peak data rate, and other such metrics may become available and could be incorporated in a richer measure of availability.

If one were to focus on a single metric to assess broadband availability, the L2 metric would be the one we would recommend.

#### 4.2.1. "L3": BB Adoption on Track

This purpose of this and the following "grade" are to provide forward-looking metrics and metrics that will be useful over time. While the first three metrics focus on ensuring universal access to a basic level of infrastructure, these last two metrics focus on ensuring that broadband markets remain on track.

The L3 metric is intended to ensure that adoption rates/service quality are at or above (and for prices, at or below) the national average, or to measure within-state disparities, at or above (below) the state average.

To look at these static comparisons, data would be needed on such things as service penetration rates or $/Mbps/month. End-user surveys may provide the best way to collect such data. Conducting statistically valid surveys to collect such data will require significant resources.

In addition to considering static comparisons, the L3 metric should also look at trends (year-on-year changes). One way to implement these would be to code as “Green” improvements (unambiguous – all significant metrics improve), “Red” as deterioration, and “Yellow” as mixed or uncertain. Such trend-based metrics may be computed using multiple years for the L0-L2 metrics.

### 4.2.2. “L4”: BB is Best in Class

The final class of measurements or “grade” to consider focuses on the performance of broadband services at the higher-end of the service continuum. In addition to ensuring universal accessibility to an appropriate minimal standard of broadband services, policymakers need to be concerned with the performance of broadband in more advanced markets. This will contribute to Massachusetts competitiveness and the ensure the health of key economic sectors that are IT-dependent like healthcare and high-technology.

It will be important to ensure that next generation broadband services such as FTTH and 3G (mobile broadband) are progressing in Massachusetts at rates that are competitive with the progress in other leading peer states (e.g., New York, California, North Carolina).

Metrics that track Massachusetts’s performance with respect to adoption, quality, competitive choices, and pricing relative to broadband leaders will be important also.

This class of metrics may include additional measures to track such things as long-haul fiber capacity or specialty requirements of key sectors against peer states.

Over the next year or two, much better federal and state data ought to become available that will enable formulation of a rich set of metrics.

In the future, we expect that the L3 and L4 metrics will be much more important in assessing the health of broadband services. However, at the current stage of data availability, the role of these metrics in the broadband report card is as placeholders and it is not our expectation that they would be populated in the first application of the report card.

Our proposal for a community broadband report card is summarized in the following figure:

| Goal | Title | Description |
|---|---|---|
| | | ***Achieving Ubiquitous Availability*** |
| L0 | Town government is on-line | • Q01: Does the town hall have BB available for its internal use?<br>• Q02: Is there a municipal intranet linking (most) government offices, libraries, schools, hospitals, police/fire?<br>• Q03: Does the town maintain an active municipal website? |
| L1 | Public access BB is available | • Q11: There are 1 or more public access terminals with BB access in the community?<br>• Q12: Public schools have BB access for educational (and internal) use?<br>• Q13: Public libraries have public access terminals with BB access? |
| L2 | Ubiquitous BB available | • Q21: Are 1B broadband or better ("1B+") services available to 95% or more of the homes? (Basic availability)<br>• Q22: Are 1B+ available from more than one facilities-based provider to 60% or more of the homes? (Competition and consumer choice)<br>• Q23: Are 2B+ available to 60% of homes? (Quality of service). This metric may be expanded to include other measures of service quality such as customer complaints or outages.<br>• Q24: Is price for 1B service at or below state average? |
| | | ***Keeping Broadband on track*** |
| L3 | BB adoption on track | • BB adoption rates are on par with national average.<br>• $/Mbps/month for average, best, and entry service on par with national averages<br>• Within state differences on par with peer states. |
| L4 | BB is best in class | • BB availability and adoption rates for higher quality BB services (2B, 3B, 4B) are on par with national averages.<br>• Within state differences on par with peer states |

# 5. Data on Broadband Availability[43]

To collect the data to test the various broadband metrics, we employed a multi-tiered strategy. First, we collected as much third-party data as possible from public data sources and from sources within the state we were put in touch with through the MTC/JAII. This included reviewing FCC and MTC data, as well as a study prepared internally by the MTC/JAII.

Wherever possible, we collected such data on a statewide basis and on as granular a basis as feasible. This varied by type of data source, as discussed in the following section where we present our results.

We also developed software tools to crawl service provider websites to identify a (pseudo) randomized sample of household locations where broadband service is available. Of the approaches employed, this is the one that proved the most interesting and potentially useful for future data collection efforts.

However, the overarching conclusions from these data collection efforts were as follows:

- Existing publicly available data sources are inadequate to support an appropriate assessment of broadband availability in Massachusetts.
- GIS-based maps (using service provider data) and large-sample (statistically valid) end-user surveys will be needed to assess broadband. [Detailed specification or implementation of these were beyond the scope of this project.]
- Significant resources are needed to support appropriate broadband metrics efforts in Massachusetts: on the order of $100k/year for even a fairly minimalist effort.

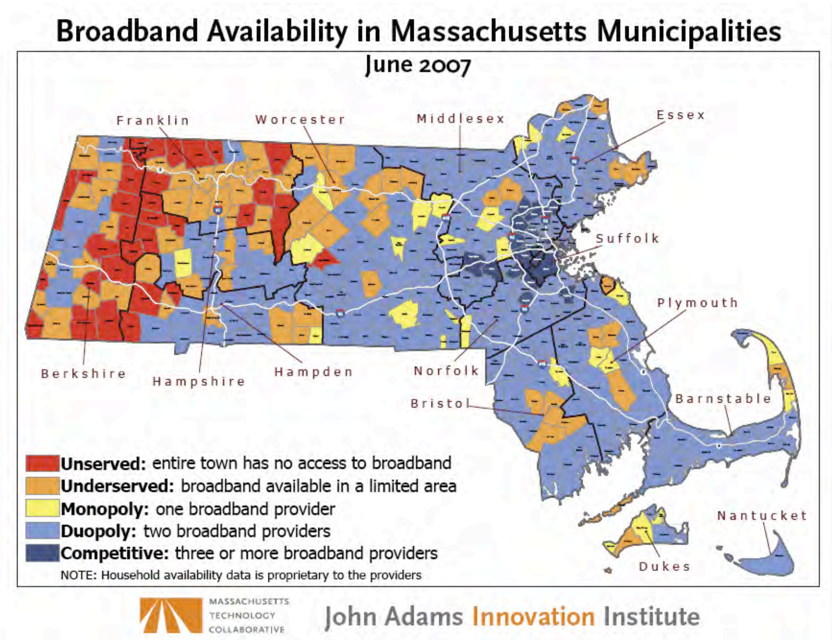
The various sources consulted are discussed in the following sub-sections.

## 5.1. MTC/JAII Data

During the first half of 2007, the MTC/JAII conducted an internal study to assess the level of broadband availability in Massachusetts on a per-town basis (for each of the 351 communities in Massachusetts). That analysis resulted in the following color-coded map:[44]

[43] The data analysis reported in this chapter is based on primary collection and analysis efforts by Tony Smith-Grieco to show the sorts of things that are possible with publicly available third-party data. This analysis shows the limited value of such data and highlights the need for better service-provider data and end-user survey data in order to appropriately assess the status of broadband services within the state.

[44] This map is available at http://www.mtpc.org/broadband/map.pdf.




The underlying data on which this report was based is from before January 2007. This data may also be presented in tabular form as follows:

**Broadband Availability in Massachusetts**[45]

| | No. Towns | HH Served | Shr HH Served |
|---|---|---|---|
| **Category of Community** | | | |
| Unserved (no DSL or Cable) | 33 | 11,725 | 0.5% |
| Underserved | 60 | 207,156 | 8.3% |
| Monopoly | 25 | 81,459 | 3.3% |
| Duopoly | 216 | 1,701,928 | 68.0% |
| Competitive | 17 | 499,521 | 20.0% |
| Total | 351 | 2,501,789 | 100.0% |
| | | | |
| DSL Communities (incl. FIOS) | 282 | 2,390,231 | 95.5% |
| No DSL | 69 | 111,558 | 4.5% |
| | | | |
| FTTH Communities | | | |
| Verizon FIOS | 71 | 581,675 | 23.3% |
| None | 280 | 1,920,114 | 76.7% |
| | | | |
| Cable Franchisees | | | |
| Charter Communications | 53 | 270,361 | 10.8% |
| Comcast Cable | 237 | 2,143,484 | 85.7% |
| Cox Communications | 1 | 1,005 | 0.0% |

[45] This table was created using the data from the spreadsheet file provided by MTC/JAII analyst Adele Burnes.

| | | | |
|---|---|---|---|
| Russell Municipal Cable TV | 1 | 658 | 0.0% |
| Shrewsbury's Community Cablevision | 1 | 13,173 | 0.5% |
| Time Warner Cable | 15 | 55,773 | 2.2% |
| None | 43 | 17,335 | 0.7% |
| Total | 351 | 2,501,789 | 100.0% |
| | | | |
| Cable Overbuilders | | | |
| BELD | 1 | 12,964 | 0.5% |
| RCN | 16 | 486,557 | 19.4% |
| None | 334 | 2,002,268 | 80.0% |

As of this writing, this analysis still represents the best composite picture of the variation in broadband availability across all towns in Massachusetts, although progress continues to be made. This data suggests that the share of households that were either unserved (no broadband available in the community) or underserved (broadband available to only part of community) was less than 10% in January 2007 and that the majority of these communities were clustered in the western part of the state. However, it should be noted that this data likely overstates the availability of broadband coverage.

## 5.2. Census Consumer Expenditure Data

The U.S. Census Bureau does not collect data about Internet access as part of its regular surveys. However, the October 2003 Current Population Survey (CPS) included a special supplement with questions on computer and Internet usage in households. At that time, about 58% of the 2,577 Massachusetts households surveyed had Internet access[46], compared to about 55% for the nation as a whole. It should be noted that this data focused on Internet access, rather than broadband access and so its relevance to ascertaining the availability of broadband services is limited.

## 5.3. FCC data

The FCC produces bi-annual reports of broadband connectivity, where broadband is defined as connections with 200 Kbps or greater bandwidth in *either* direction (upload or download). The most geographically detailed data is a table of the number of broadband providers in each zip code. This data is somewhat limited, as a provider may be counted for a zip code when it only serves a single subscriber in that zip code. There is no way to know if the provider serves the entire zip code, or just a very small portion of it. In addition, it is likely that many of the providers are focused on the business market and not the residential market. Thus the data gives a somewhat optimistic picture of the level of competition for Internet service. However, it may still be useful to look at the patterns in broadband service indicated by this data. The following tables summarize the important information provided in the FCC data, and the accompanying map illustrates the variability in high-speed Internet service across the state. There are some

[46] http://www.census.gov/population/www/socdemo/computer/2003.html, Table 1B, "Presence of a Computer and the Internet for Households, by State: October 2003". 64% of households surveyed had a computer.

areas where no Internet service providers are in operation, and a few areas where there are as many as 17 different providers. The bulk of zip codes have between 5 and 7 providers reporting. About a sixth of zip codes have fewer than this number of providers, and the remaining third of zip codes have more.

The map shows that even at a general level, there is a definite difference between the number of broadband service providers available in western vs. eastern Massachusetts. It also shows some correlation with the broadband status determined by JAII. Many of the JAII-labeled underserved and unserved towns have a lower number of providers (5 or 6), as do some of the monopoly and duopoly towns. However, as an illustration of the problems with this FCC data, it shows a higher number of providers (9 or 10) in several towns in the far west, which JAII labeled underserved or unserved.

While this data is not detailed enough to draw more specific conclusions, it does give an illustration of the digital divide in the Commonwealth.

### Table 1. FCC broadband statistics for Massachusetts (as of June 30, 2007)[47]

| | |
|---|---|
| 2006 estimated total number of households (from the U.S. Census[48]) | 2,708,986 |
| Residential high-speed lines | 1,705,007 |
| Business high-speed lines | 955,494 |
| Total residential and business | 2,660,501 |
| # of high-speed lines by technology (both business and residential): | |
| DSL | Confidential (not reported) |
| Cable modem | 1,088,170 |
| Traditional wireline | 16,986 |
| Fiber, satellite, fixed and mobile wireless | Confidential |

[47] FCC Report, "High-Speed Services for Internet Access: Status as of June 30, 2007", http://www.fcc.gov/wcb/iatd/comp.html, published March 2008.

[48] U.S. Census Population Estimates Program, http://www.census.gov/popest/estimates.php

## Table 2. Adjusted[49] FCC statistics on broadband providers in Massachusetts (as of June 30, 2007)

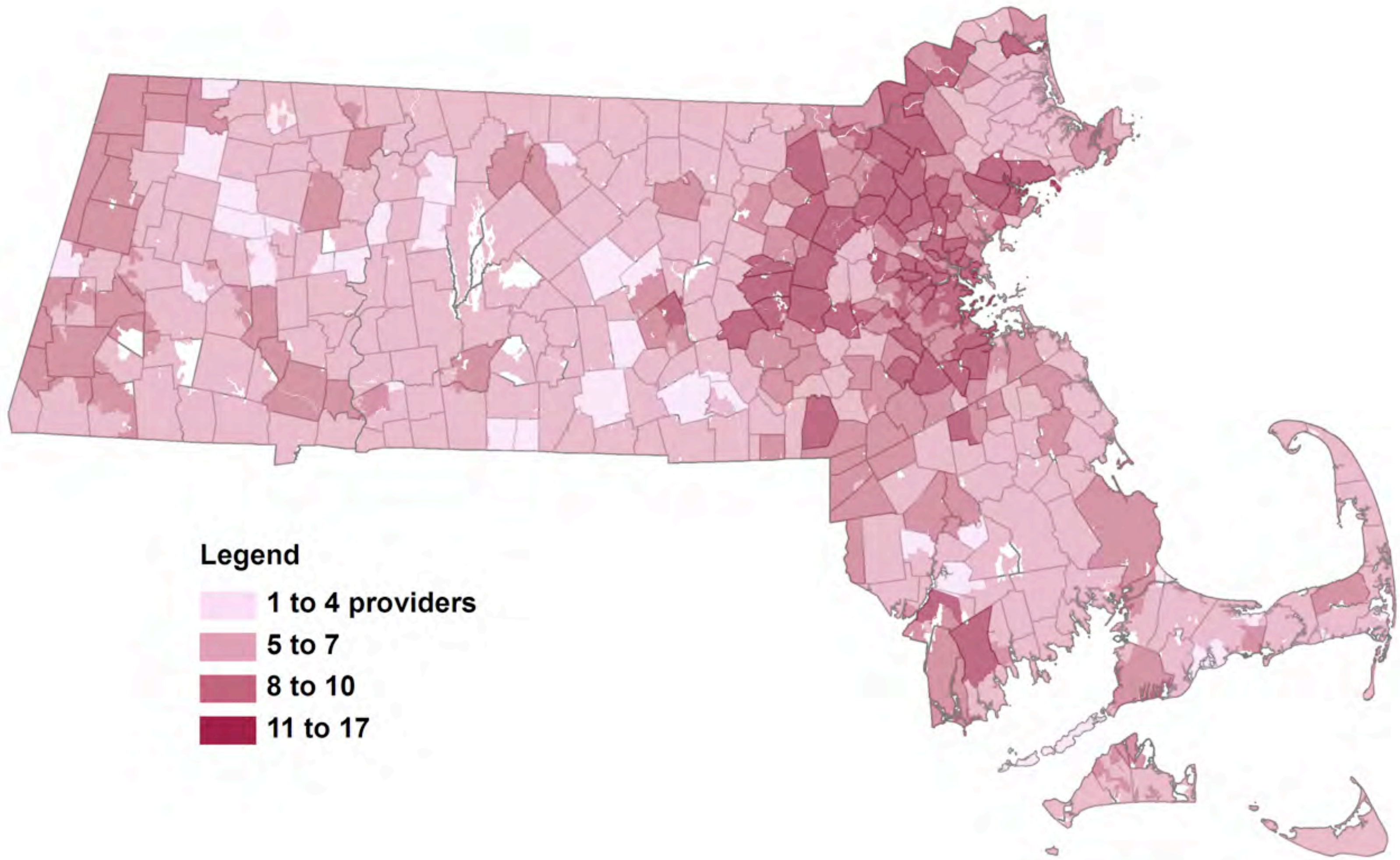
| | |
|---|---|
| DSL providers (both ADSL and SDSL) | 24 |
| Traditional wireline providers | 17 |
| Cable Internet providers | 7 |
| Fiber providers | 6 |
| Satellite, fixed wireless and mobile wireless providers | 1 to 3 (in each category)[50] |
| Other providers (including broadband over power line) | 0 |
| Total number of Internet providers[51] | 38 |
| % of zipcodes with no providers | 6 |
| % with 1 to 4 providers | 9 |
| % with 5 to 7 providers | 47 |
| % with 8 to 10 providers | 21 |
| % with 11 to 13 providers | 14 |
| % with 14 to 17 providers | 3 |
| Median number of providers in a zip code | 7 |

[49] The "% of zipcodes with N providers" numbers do not match the numbers given in the FCC report, because the FCC does not count the zip codes which do not have any service providers. This number must be estimated from other sources. The method of estimation is explained in the appendix.

[50] Exact number withheld to protect firm confidentiality.

[51] Some companies may provide services using more than one kind of technology, so this number does not equal the sum of the other numbers.

**Figure 2. Number of high-speed Internet service providers in Massachusetts zip codes[52], according to FCC data for June 30, 2007**




[52] Zip codes boundaries come from the GeoLytics Estimates Professional 2007 product.

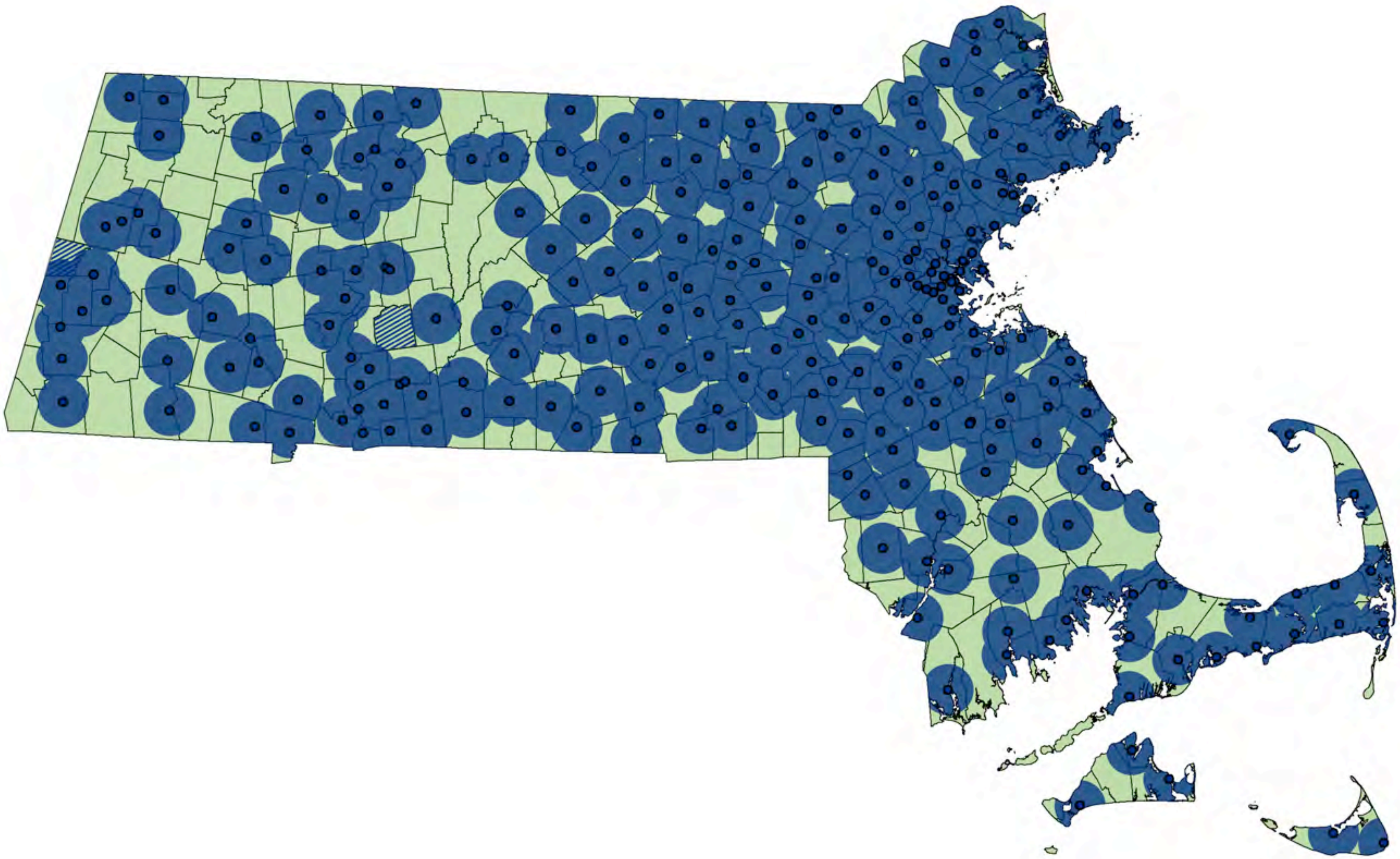
## 5.4. Theoretical DSL coverage gaps based on wirecenter locations

Based on the locations of wire centers, it is possible to determine which areas could not possibly receive DSL service, because they are simply too far from a wire center. Households that are more than 18,000 feet from a wire center can not receive DSL.

Using a list of Verizon wire center locations, from the Telcordia *Local Exchange Routing Guide,* a map was generated showing the *best-case, most optimistic* coverage areas for DSL, by drawing circles with a 18,000-foot radius around each of the wire centers. *Just because a household lies within the dark blue area does* ***not*** *mean it can receive DSL service.* However, it can be assumed that households in the green areas can **not** receive DSL. This estimate is a rough-order-of-magnitude at best: cabling must follow street patterns rather than birds-flight, but it is also possible to extend coverage by deploying remote terminals and other equipment not located in the central office.

There are two towns, Richmond and Granby, which have wire centers owned by other phone companies, for which we do not know the locations. Those towns are marked with cross-hatched blue lines.

**Figure 3. Verizon wire centers and DSL coverage gaps**



## 5.5. Provider websites – sampling specific addresses

For the selected set of towns, we have also chosen a set of addresses at random and checked Internet availability at those addresses on the Comcast and Verizon websites.

For each town, we collected 100 residential entries from a public White Pages website, whitepages.com. A search was done for all entries in each town having a last name beginning with the letter L, and the first 100 results were used. Most entries included both phone numbers and addresses, although in some cases, either the address or the phone number was not listed. An automated "web scraping" tool[53] was then used to enter the addresses into the Comcast website and the phone numbers into the Verizon website. (Some addresses were P.O. boxes and therefore could not be entered into the Comcast website. Also, the Verizon website reported that some of the phone numbers were not Verizon landline numbers, in which case broadband availability could not be determined.)

For all but one of the towns, the Comcast website reported availability at all of the sampled points. The exception was Harvard, which is served by Charter Communications, not Comcast (according to JAII's June 2007 report).

On the other hand, the Verizon website reported much more variability in broadband service. The following table shows the number of points for which service was reported as available, not available, and unknown. "Available" includes locations where either DSL or FiOS (Verizon's fiber-to-the-home offering) are available. "Unknown" means that the website displayed the message "Sorry, we are unable to determine if Verizon High Speed Internet is available at your location" and suggests calling the local Verizon office to determine availability. The "% available" was calculated by dividing the "available" number into the sum of the "available" and "not available" numbers (thus ignoring the "unknown" points).

This is intended not as a final answer to the question of underservedness, but as an illustration of one possible way to collect some additional information. This data can be collected with off-the-shelf, free software tools.

Following the table are maps of the towns, with color-coded dots indicating locations of availability and non-availability. It is interesting to note that there are points of non-availability very close to points of availability.

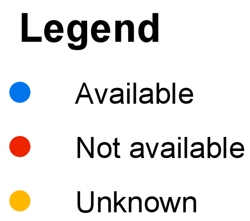
[53] The free *iMacros* extension for the Mozilla Firefox browser, version 6.0.3.4, was used. It was designed by iOpus Software: http://www.iopus.com/imacros/

**Table 3. Verizon broadband availability at selected locations, according to the Verizon website**

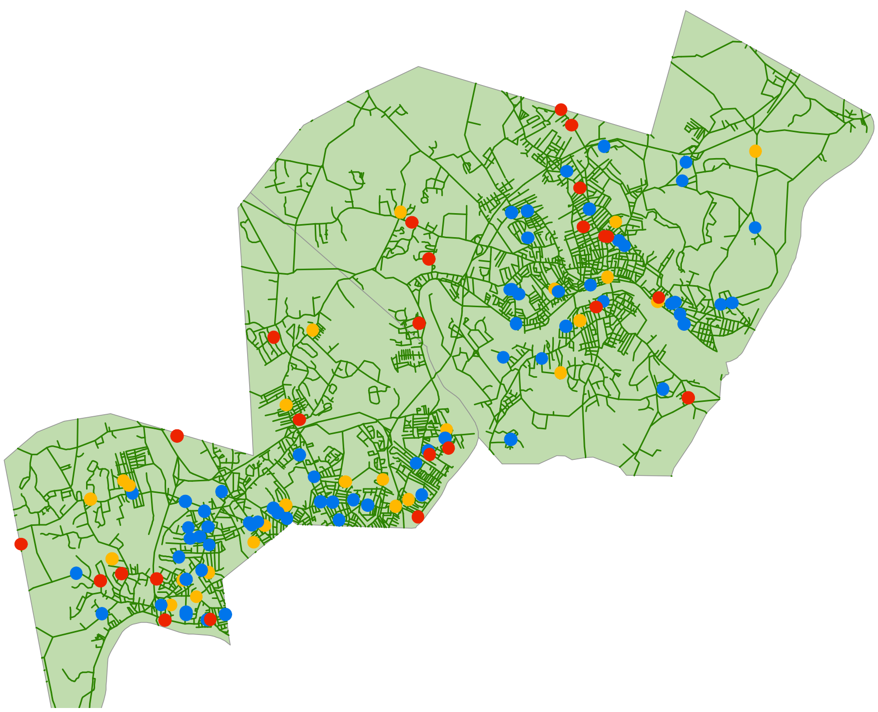
| Town | JAII broadband status[54] | Number of sampled households where broadband service is… Available | Not available | Unknown | Estimated % available |
|---|---|---|---|---|---|
| Concord | Monopoly (no DSL) | 51 | 22 | 11 | **70** |
| Eastham | Monopoly (no DSL) | 29 | 35 | 13 | **45** |
| Harvard | Monopoly (no DSL) | 38 | 27 | 15 | **58** |
| Haverhill | Duopoly | 35 | 13 | 8 | **73** |
| Lexington | Competitive (Comcast, FiOS, RCN) | 29 | 13 | 31 | **69** |
| Methuen | Monopoly (no DSL) | 37 | 14 | 18 | **73** |
| Orleans | Duopoly | 42 | 14 | 16 | **75** |

[54] The broadband status of the town according to JAII's June 2007 report.

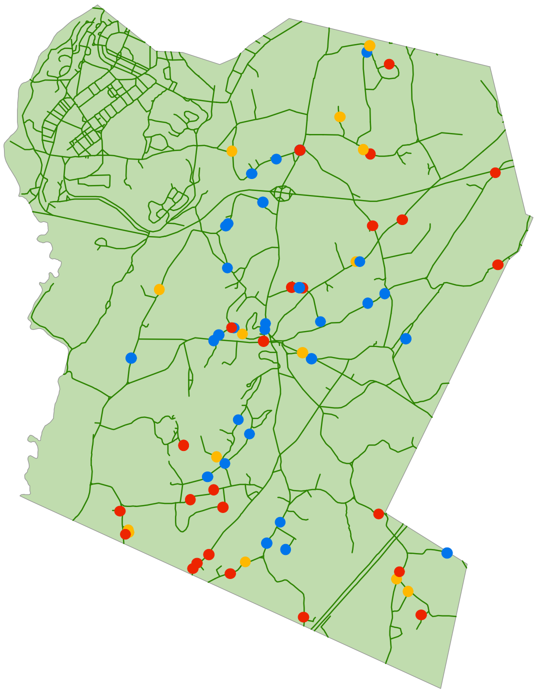
**Figure 4. Maps of Verizon broadband availability in selected towns**



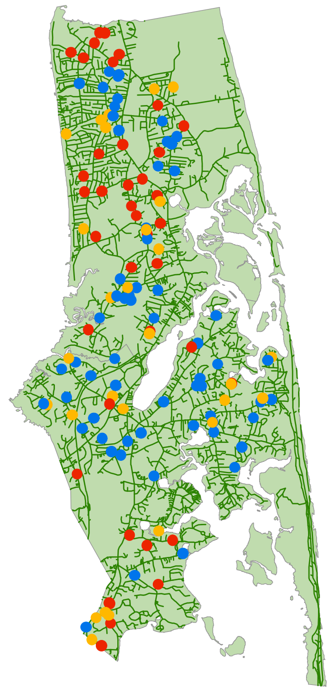

### Haverhill and Methuen



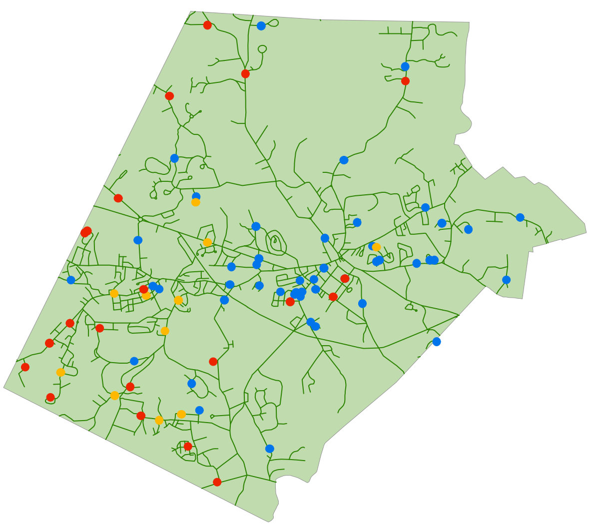
**Harvard**



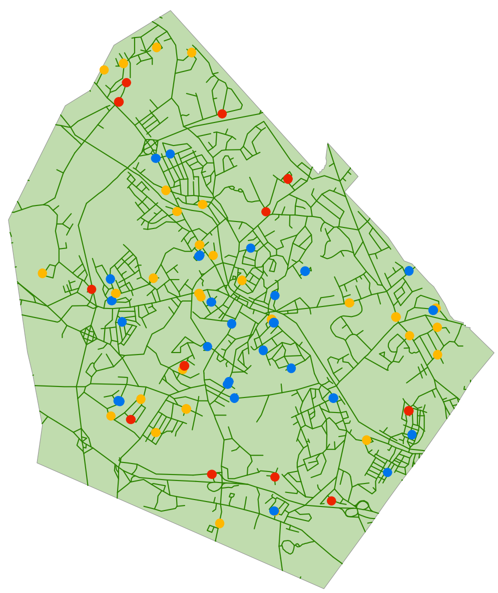
**Eastham and Orleans**



**Concord**



**Lexington**



## 5.6. Speedtest.net bandwidth test data

The website speedtest.net offers a free bandwidth test for individuals to measure the speed of their broadband connections. The operators of the site, a company named Ookla, shared the data they collected for Massachusetts from July of 2006 through January of 2008, a period of about eighteen months. Because this test is entirely voluntary, it is difficult to draw reliable conclusions about broadband availability from the data, but we believe it might be useful as a crosscheck on other data. We first discuss some of the complications in interpreting the data, and then present the results of a preliminary analysis.

The first difficulty is that while it is known how many speed tests have been performed, it is difficult to know if those tests represent distinct households, or just a single household doing the test repeatedly. The site does not collect any data from users which could be used to identify particular users or households. This is consistent with sound privacy policies, but does complicate the analysis of the data for our purposes. All that is known for certain is the user's IP (Internet Protocol) address, which is a set of numbers that together identify a computer for the purpose of routing data across the Internet *at one moment in time*. However, this IP address is not a unique permanent identifier for the computer or household; there is no guarantee that a household will have the same IP address today that it had yesterday or will have tomorrow. ISPs are free to assign IP addresses to customers more or less as they please, and may use complex proprietary internal algorithms to do so. For some ISPs, IP addresses are a scarce resource which needs to be rationed, whereas others may not be so concerned. This is not an issue particular to Speedtest.net, rather it is a general issue that any automated broadband measurement system will face.

For similar reasons, it is difficult to determine with certainty where users are located. The IP address contains no geographical information; it only identifies the owner of the network (the ISP) from which the user is connecting. Short of asking users to enter their location manually, the only way to geo-locate a particular IP address is to ask the ISP. There are however various independent software vendors which have made these contacts with ISPs and sell databases which predict a location based on an IP address. Such products are not 100% accurate, but they can provide some useful information. For example, they are used by online advertising services to target ads to people in particular geographical areas. Speedtest.net uses one such product, GeoIP, from the company MaxMind based in Boston.

A further complication is that the city/town names provided by GeoIP are the town names used by the US Postal Service in mailing addresses, which do not always correspond with the town names officially recognized by the state. For example, there is a significant amount of data for the towns of "Newton Center," "Brighton," and "Wellesley Hills," which are not official state-recognized names, but are used by the USPS. In total, there are 494 distinct "towns" in the Speedtest data set for Massachusetts (compared to the 351 official municipalities recognized by the state). The data set does include latitude and longitude coordinates for the towns, which we used for GIS mapping.

In addition, the measured upload and download speeds will be somewhat dependent on the proximity of the user's computer to the servers running the speed test software. Speedtest operates a network of servers around the world, and some users will be closer to those servers than others. In addition, momentary congestion, bottlenecks, or other problems in the user's Internet connection can affect the measured speed. The speed test only measures the bandwidth at one moment in time, while the user is waiting; it does not do repeated tests over the course of days or weeks.

Despite these caveats, we did some initial analysis of the data to see how useful it might be. The data set we used was aggregated on the basis of town and ISP. In other words, for each ISP there is an average download and upload speed for each town, as well as the number of distinct IP addresses which had measurements for that town. This data set did not include the number of speed tests or their dates. Ookla has indicated they would be willing to provide the disaggregated (raw) data for a fee, including data on each individual speed test, which might be useful for more advanced analysis.

The following table gives an example of the data. It includes the 10 pairs of cities and ISPs with the largest numbers of distinct IP addresses, and the average download and upload speeds in megabits (Mbps) across all of those IP addresses. Most of this data relates to residential broadband service, but one row gives data for MIT, which is identified as its own ISP.

| City | ISP | # of distinct IP addresses | Average download speed (Mbps) | Average upload speed (Mbps) |
|---|---|---|---|---|
| Boston | Verizon Internet Services | 6784 | 2.2 | 0.5 |
| Boston | Comcast Cable | 3084 | 9.1 | 1.3 |
| Worcester | Charter Communications | 3053 | 4.3 | 0.5 |
| Cambridge | Comcast Cable | 1910 | 8.5 | 1.1 |
| Cambridge | Massachusetts Institute of Technology | 1225 | 17.4 | 9.0 |
| Lowell | Comcast Cable | 1211 | 9.4 | 1.2 |
| Quincy | Comcast Cable | 1192 | 9.3 | 1.3 |
| Waltham | Verizon Internet Services | 1110 | 4.7 | 1.3 |
| Andover | Verizon Internet Services | 1061 | 6.8 | 1.6 |
| Worcester | Verizon Internet Services | 1003 | 1.8 | 0.4 |

The following table lists the top 25 Internet Service Providers in Massachusetts in terms of the total number of distinct IP addresses in the Speedtest.net data set. Some of these are business providers or networks. The residential providers are marked in boldface.

| Internet Service Provider | Total # of distinct IP addresses |
|---|---|
| **Comcast Cable** | **49813** |
| **Verizon Internet Services** | **41713** |
| **Charter Communications** | **10961** |

| | |
|---|---|
| **RCN Corporation** | **2982** |
| Comcast Business Communications | 1676 |
| Massachusetts Institute of Technology | 1229 |
| **Road Runner** | **1220** |
| Cellco Partnership DBA Verizon Wireless | 1073 |
| **AT&T WorldNet Services** | **937** |
| Covad Communications | 709 |
| Conversent Communications | 694 |
| Verizon Business | 595 |
| XO Communications | 586 |
| Sprint PCS | 517 |
| **Optimum Online (Cablevision Systems)** | **466** |
| Harvard University | 454 |
| **EarthLink** | **441** |
| CTCCommunications | 432 |
| University of Massachusetts | 432 |
| Paetec Communications | 423 |
| Boston University | 416 |
| **Adelphia** | **409** |
| **Shrewsbury Electric & Community Cable** | **376** |
| Sprint | 342 |
| Level 3 Communications | 340 |
| *Total* | 119,236 |
| *Total for all providers* | 128,618 |

We did some further analysis focusing just on the data for the 10 residential ISPs highlighted in the table. We calculated the total number of distinct IP addresses, from all of those residential ISPs, for each town in the data set (note again that the list of towns they use is somewhat different from the official list of municipalities). The results are as follows:

- 281 towns have fewer than 100 total distinct IPs,
- 156 towns have between 100 and 500,
- 42 towns have between 500 to 1000, and
- 15 towns have over 1000.

As expected, most of the latter 15 towns are in the eastern half of the state. The furthest west is Chicopee, and after that the furthest west is Worcester. At the end of this section is a table listing the top 50 towns in the Speedtest data set, in terms of the total number of IPs from residential ISPs. The only towns west of the Springfield-Chicopee-Northampton area with more than 100 total IP addresses are Pittsfield (474), North Adams (156), Huntington (171), and Worthington (130). We cite these numbers here for purely illustrative purposes, since this does not imply a statistically valid test.

As an interesting point of comparison with the MTC/JAII data, several towns (listed in the following table) identified as Underserved or Unserved by MTC/JAII have more than 100 IPs in the data set.

| Town | Total # of IP addresses |
|---|---|
| Carver | 112 |
| Cohasset | 116 |
| Worthington | 130 |
| Southampton | 166 |
| Greenfield | 180 |
| Fitchburg | 299 |
| Pembroke | 304 |
| Gloucester | 312 |
| Billerica | 330 |
| South Hadley | 432 |
| West Springfield | 568 |
| Fall River | 865 |

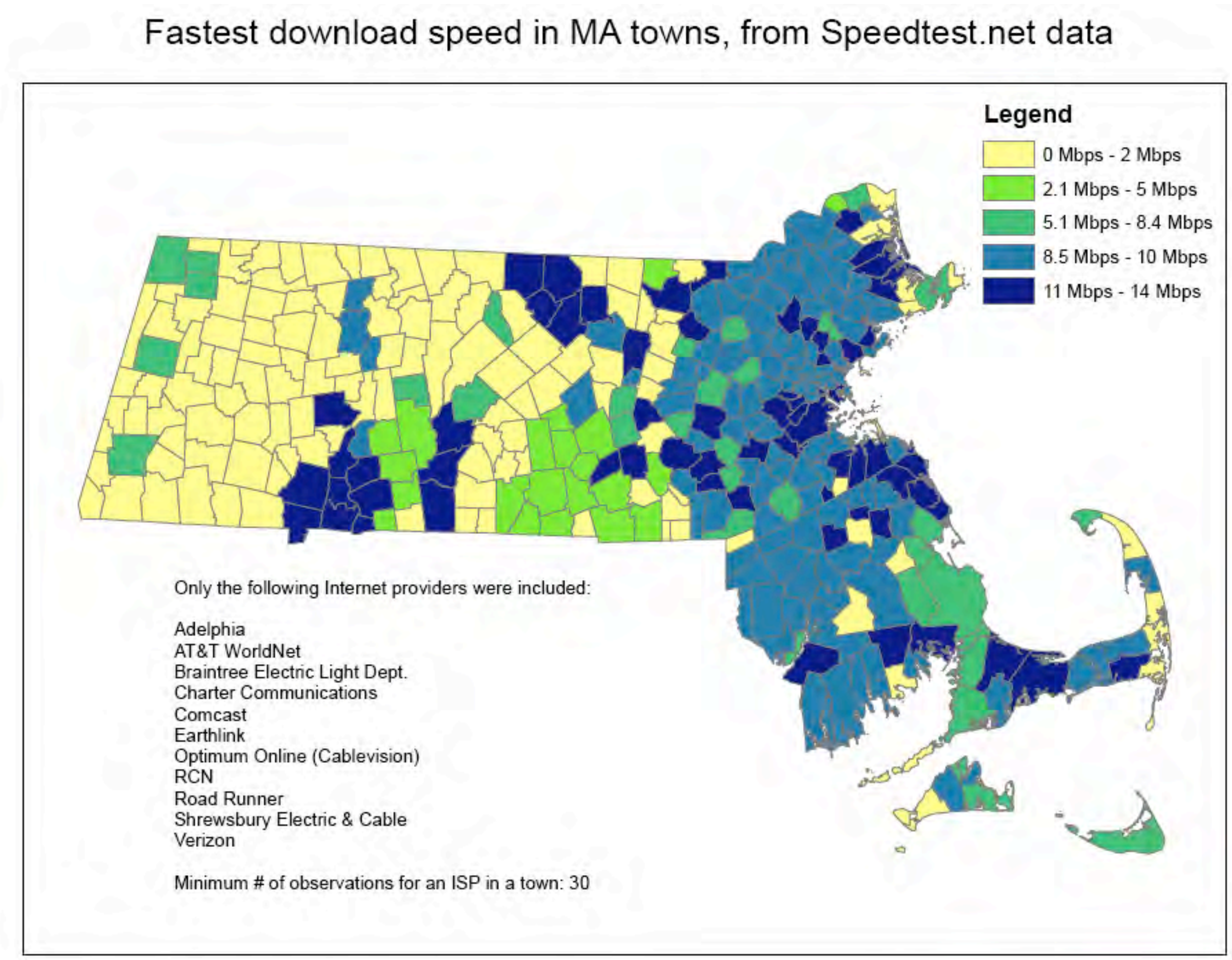
Also, all but 3 of the top 50 towns listed in the table below are identified as Competitive or Duopoly by MTC/JAII. The 3 exceptions are Fall River (Underserved), West Springfield (Underserved), and Weston (Monopoly).

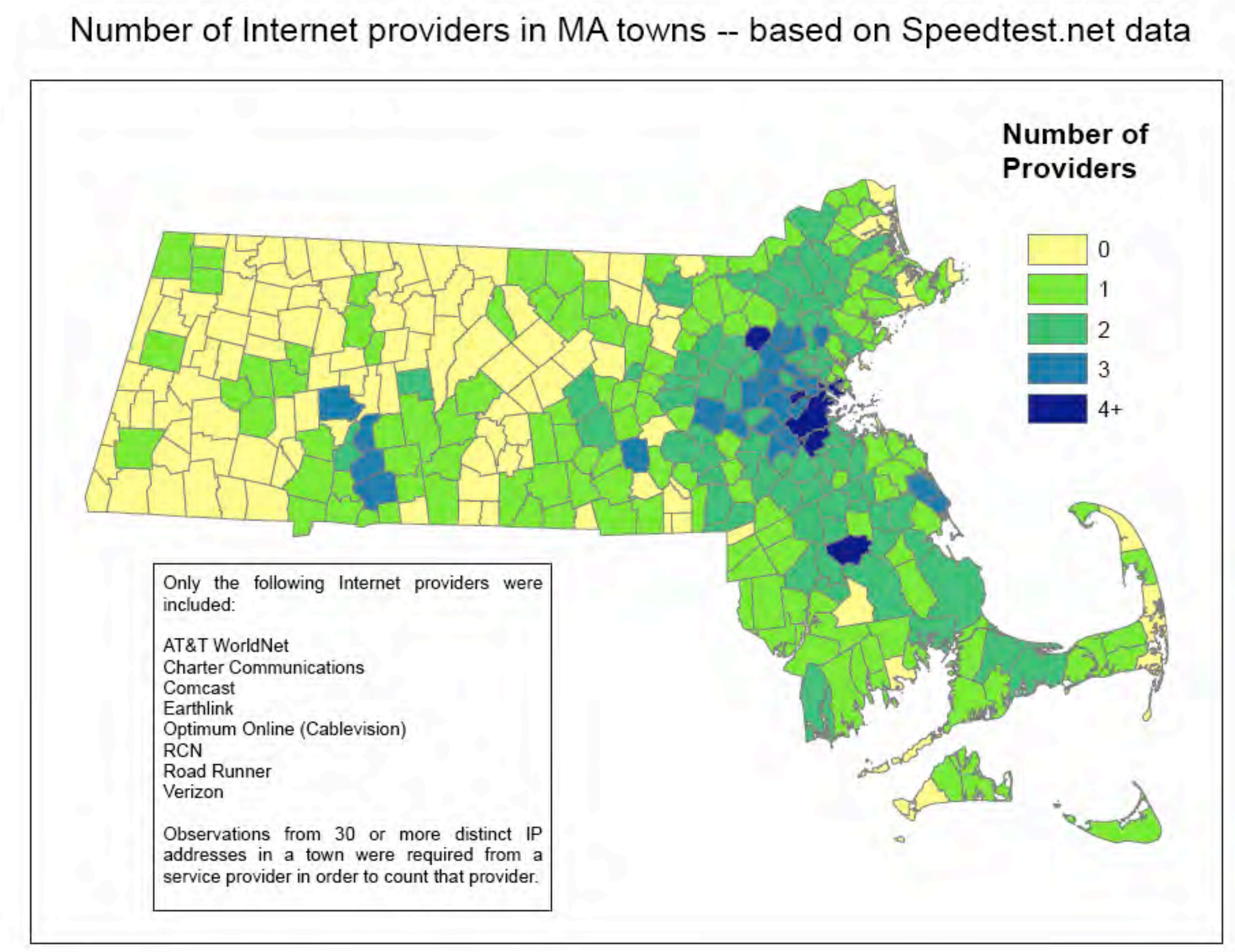
We have also mapped the Speedtest data. The first map shows the number of ISPs in each town which had data from more than 30 distinct IP addresses. In other words, if both Verizon and Comcast had data for more than 30 IP addresses in a town, the town is marked with a darker green color; whereas if none of the selected ISPs had more than 30 IP addresses, it is yellow. The second map shows the download speed for each town, again counting only ISPs with more than 30 IP addresses. If more than one ISP had 30 IP addresses in a town, the fastest download speed was used, so the download speed indicated *may not* have been experienced by a majority of users in that town. It is more indicative of the "best case" download speed. Again, the general pattern of the results is not surprising: the choice of ISPs and download speeds are both greater in the eastern part of the state than the western part, with the exception of the Springfield metro area. In addition, in the east, the choice of ISPs and download speeds generally decrease as one moves further away from central Boston.

Our general conclusion about the Speedtest data is that it is insufficient to use on its own as a metric of broadband service, but may provide a useful complement to other data sets. For example, if other data sources indicate that a particular town does not have broadband service, yet Speedtest reports a large number of speed tests, it may be worthwhile to do further investigation of the other data sources.

Top 50 towns in the Speedtest data set, according to the total number of IP addresses from residential providers.

| Town | Total # of Speedtest IPs |
|---|---|
| Boston | 10213 |
| Worcester | 4095 |
| Cambridge | 2745 |
| Waltham | 2163 |
| Somerville | 1657 |
| Andover | 1601 |
| Quincy | 1436 |
| Lowell | 1400 |
| Framingham | 1276 |
| Marlborough | 1149 |
| Chicopee | 1053 |
| Acton | 1034 |
| Woburn | 1016 |
| Malden | 1014 |
| Brookline | 1009 |
| Easthampton | 972 |
| Springfield | 969 |
| Natick | 965 |
| Westborough | 960 |
| Arlington | 959 |
| Lynn | 923 |
| Northampton | 913 |
| Brockton | 888 |
| Newton Center | 876 |
| Fall River | 865 |
| Lawrence | 843 |
| Brighton | 841 |
| Milton | 813 |
| Taunton | 783 |
| Dedham | 773 |
| Bedford | 753 |
| Plymouth | 735 |
| Medford | 724 |
| Bridgewater | 719 |
| Needham | 699 |
| North Andover | 695 |
| New Bedford | 686 |
| Weston | 678 |
| Wellesley Hills | 673 |
| Lexington | 662 |
| Tewksbury | 644 |
| Haverhill | 642 |
| Westwood | 616 |
| Chelmsford | 601 |
| Holyoke | 593 |
| Pepperell | 592 |
| Middleboro | 580 |
| West Springfield | 568 |
| Melrose | 554 |
| Wayland | 544 |


Fastest download speed in MA towns, from Speedtest.net data
Legend
0 Mbps - 2 Mbps
2.1 Mbps - 5 Mbps
5.1 Mbps - 8.4 Mbps
8.5 Mbps - 10 Mbps
11 Mbps - 14 Mbps
Only the following Internet providers were included:
Adelphia
AT&T WorldNet
Braintree Electric Light Dept.
Charter Communications
Comcast
Earthlink
Optimum Online (Cablevision)
RCN
Road Runner
Shrewsbury Electric & Cable
Verizon
Minimum # of observations for an ISP in a town: 30


Number of Internet providers in MA towns -- based on Speedtest.net data
Number of Providers
0
1
2
3
4+
Only the following Internet providers were included:
AT&T WorldNet
Charter Communications
Comcast
Earthlink
Optimum Online (Cablevision)
RCN
Road Runner
Verizon
Observations from 30 or more distinct IP addresses in a town were required from a service provider in order to count that provider.

# 6. Appendices

6.1 Appendix A – Notes about data sources

6.2 Appendix B – Town-level survey of broadband infrastructure

6.3 Appendix C -- Technical programming code notes

## 6.1. Appendix A – Notes about data sources

The following software packages and data sets were used in the making of this report:

| | |
|---|---|
| GIS software | ArcGIS 9.2 Build 1380, ESRI |
| Geocoding and street maps | StreetMap USA 2004, included with ArcGIS<br><br>("Geocoding" refers to the process of converting street addresses to latitude/longitude coordinates so that they can be mapped. This was necessary to map the broadband availability data gathered from the Verizon website.) |
| Zip code boundaries | Estimates Professional 2007, GeoLytics, Inc.<br><br>http://geolytics.com/USCensus,Estimates-Professional,Products.asp |
| Massachusetts town boundaries | Community Boundaries (Towns) from Survey Points, MassGIS<br><br>http://www.mass.gov/mgis/townssurvey.htm |
| Wire center locations | Local Exchange Routing Guide (LERG), Telcordia Technologies, Inc.<br><br>http://telcordia.com/products_services/trainfo/catalog_details.html#LERG<br><br>This database contains VH coordinates for the wire centers; those coordinates were converted to latitude/longitude using the open-source Perl module **Geo-Coordinates-VandH-1.10,** downloadable at<br><br>http://search.cpan.org/~ptimmins/Geo-Coordinates-VandH-1.10/VandH.pm |

### 6.1.1. FCC zip code data

The FCC does not report zip codes with no Internet service providers. In order to identify those zip codes, we used the GeoLytics Estimates Professional 2007 data set, which has a list of all of the zip codes in Massachusetts. This file lists 519 total zip codes in Massachusetts, whereas only 496 zip codes appeared in the FCC data. 7 zip codes in the FCC list did not appear in the Geolytics list. Thus we estimate a total number of 526 zip codes in the state, 30 of which have no high-speed Internet providers.

The 7 zip codes not included in the Geolytics file are as follows, along with the corresponding town names as given by the US Postal Service website[55]. The FCC data file listed all of these zip codes as having 5 or fewer Internet service providers.

| | |
|---|---|
| 01063 | Northampton (Smith) |
| 01343 | Drury / Charlemont |
| 01354 | Gill (Northfield Mt. Hermon) |
| 01434 | Devens / Ayer |
| 02153 | Medford (Tufts) |
| 02325 | Bridgewater (Bridgewater St. College) |
| 02357 | North Easton (Stonehill College) |

[55] The "Find All Cities in a ZIP Code" service was used: http://zip4.usps.com/zip4/citytown_zip.jsp

## 6.2. Appendix B – Town-level survey of broadband infrastructure

The following is a survey instrument that could be used to collect information about the level of broadband service in a town or community, in order to determine its level of "underservedness".

Introduction

1. **Please enter your name.**

2. **For which town/municipality/community are you completing this survey?**

3. **Please enter the street address of your town hall, including the zip code.**

4. **Please enter a phone number where you may be reached.**

5. **Please enter your email address.**

6. **Community information. What are the major employers and major industry sectors in your community? If particular companies have large facilities in the area, please name them. Also please tell us the major types of industries (for example, manufacturing, agriculture, or health care) that employ people in your community.**

7. **What is the approximate median household annual income in your community? (in dollars)**

8. **Approximately what percentage of people in your community are below the poverty level?**

Town hall/government Internet usage

9. **For each government activity/building type, please indicate if they have broadband:**

| | Yes, they have broadband service | No broadband available | Don't know if broadband available | Broadband available, but do not subscribe | Broadband available, but don't know if they subscribe |
|---|---|---|---|---|---|
| Town hall | | | | | |
| Police department | | | | | |
| Fire department | | | | | |
| Hospital | | | | | |

Other or Comments

**10. Please indicate what kinds of Internet connections these government services have. You may check more than one box for each.**

| | DSL | Cable modem | T1 | Fiber | DS3 | OC1, OC3, etc. | Ethernet | Satellite | Wireless |
|---|---|---|---|---|---|---|---|---|---|
| Town hall | | | | | | | | | |
| Police department | | | | | | | | | |
| Fire department | | | | | | | | | |
| Hospital | | | | | | | | | |

Other type of connection (please specify)

**11. What company(s) provide Internet service for these facilities?**

| | |
|---|---|
| Town hall | |
| Police department | |
| Fire department | |
| Hospital | |

**12. Approximately how much do each of these facilities pay each month for their Internet service?**

| | |
|---|---|
| Town hall | |
| Police department | |
| Fire department | |
| Hospital | |

**13. Please enter the full addresses (street, city, and zip code) of any local government buildings where Internet service is not available:**

| | |
|---|---|
| Address #1 | |
| #2 | |

| | |
|---|---|
| #3 | |
| #4 | |
| #5 | |

**14. Overall, for the town government facilities which do have Internet connections, are they satisfied with the quality of their connection?**

Yes No

**15. If they are not satisfied, what are the reason(s) for dissatisfaction? (You may select more than one)**

| | |
|---|---|
| Price is too high | |
| Connection is too slow | |
| Service is unreliable | |
| Not enough choice of providers | |
| Other (please specify below) | |

Please list any other reasons for dissatisfaction:

<u>Residential Internet service providers</u>

**16. What is the dominant phone company in your area?**

Verizon
Richmond
Sentinel Tree
Taconic (now FairPoint)
Other (please specify)

**17. Which company is the dominant cable provider in your area?**

Comcast
Charter Communications
Time Warner
Other (please specify)

**18. Are there any other companies which provide residential Internet service to your community *with their own facilities or infrastructure* ? If so, please give their names below.**

**In other words, please do not include companies which offer service with lines leased from the main phone or cable company.**

Yes No

Names of additional Internet providers:

**19. Please complete the following worksheet for the phone company, cable company and any other providers of residential Internet service in your community.**

## Internet service provider worksheet

**A. Name of this company:**

**B. This company provides Internet service to what share of the households in your community (approximately)?**

None
Less than 25%
25% to 49%
50% to 74%
75% to 99%
The entire community
Don't know

Comments

**C. If this company provides Internet service, we are interested in learning about the range of choices offered (at least in some part of the community). Please rate the service with respect to the peak rated download speed.**

| | Less than 500 Kbps | 500 Kbps – 999 Kbps | 1 Mbps – 4.9 Mbps | 5 Mbps – 49 Mbps | 50 Mbps or greater | Don't know |
|---|---|---|---|---|---|---|
| Entry-level / least expensive service | | | | | | |
| Average (most common) service | | | | | | |
| Premium / best service | | | | | | |

Comments

**D. If this company provides Internet service, what is the fastest Internet speed that is available to the *majority* (over 50%) of households?**

**In other words, if a small number of the company's customers can get 100 Mbps service, but the rest of the customers can only get 5 Mbps service, then please select "5 to 49 Mbps".**

**Also, for the purpose of this survey, we are concerned only with download**

**speed, not with upload speed.**

Below 500 Kbps
500 to 999 Kbps
1 to 4.9 Mbps
5 to 49 Mbps
50+ Mbps

**E. What are the installation costs and monthly prices of this company's Internet service offerings?**

| | One-time installation / setup cost | Monthly price |
|---|---|---|
| Cheapest offer | | |
| Average/most common offer | | |
| Best/most expensive offer | | |

The community as a whole

We now ask questions about availability of service in the community as a whole. We would like to find out if the providers are all serving roughly the same areas, and if people have a choice of providers. We realize it will be hard to answer these questions exactly. Please try your best to estimate the availability of service.

**20. Approximately what portion of the community as a whole can get Internet service at the following speeds? And what portion can get service from multiple providers?**

**For each box, just enter a choice from the following list:**

None
Less than 25%
25 to 49%
50 to 74%
75 to 99%
The entire area
Don't know

| | From just one provider | From two or more providers |
|---|---|---|
| 1 to 4.9 Mbps | | |
| 5 Mbps and above | | |

Comments

**21. For the next questions, we would like you to provide information about the availability of Internet service in particular areas/neighborhoods in your community. Please list at least four distinct areas/neighborhoods in your community. For example, if you were completing this survey for Boston, the area names might include "Allston", "Dorchester", "Roxbury", and "Jamaica Plain" (as well as others). However, these areas need not correspond to commonly-used neighborhood designations. It is more important that they indicate areas which have greater or lesser levels of Internet service availability.**

| | |
|---|---|
| Area #1 | |
| #2 | |
| #3 | |
| #4 | |
| #5 | |
| #6 | |
| #7 | |
| #8 | |
| #9 | |
| #10 | |

**22. Now, for each of the areas listed above, please estimate what portions of the households in each area have access to Internet services at various speeds. For each box, you may just enter one of the choices from the following list:**

None
Less than 25%
25 to 49%
50 to 74%
75 to 99%
The entire area
Don't know

| | 1 to 4.9 Mbps | 1 to 4.9 Mbps, from more than one provider | 5 Mbps and above | 5 Mbps and above, from more than one provider |
|---|---|---|---|---|
| Area #1 | | | | |
| #2 | | | | |
| #3 | | | | |
| #4 | | | | |
| #5 | | | | |
| #6 | | | | |
| #7 | | | | |
| #8 | | | | |
| #9 | | | | |
| #10 | | | | |

Comments

Public Internet access points (Hotspots)

We would now like to collect information about public Internet access points in your community.

**23. Please enter the following information about Internet access at schools, libraries and community centers in your community.**

| | Public schools | Community colleges and universities | Public libraries | Community centers |
|---|---|---|---|---|
| How many such locations are there in your community? | | | | |
| How many of them have an Internet connection? | | | | |
| How many of them provide Internet access to students and teachers? | | | | |
| How many of them provide Internet access to the general public? | | | | |

**24. If Internet service is NOT available for any of these facilities, please enter the address(es) where service is NOT available. These addresses will be used for mapping purposes, to gain a better sense of which areas do and do not have Internet access.**

**Include the full street address, city and zip code.**

**If there are many places where service is not available, you may simply describe the general areas or regions where it is not available.**

| | Name of location | Street address |
|---|---|---|
| Location #1 | | |
| #2 | | |
| #3 | | |
| #4 | | |
| #5 | | |
| #6 | | |
| #7 | | |
| #8 | | |

| | | |
|---|---|---|
| #9 | | |
| #10 | | |

**25. We would now like you to enter the address(es) of locations which provide Internet access to the public. These addresses will be used for mapping purposes, to gain a better sense of which areas do and do not have Internet access.**

**Please enter the addresses of the 10 schools, colleges, universities or libraries which have the most publicly-accessible computers with Internet. Count only computers which are available to the general public, not just students.**

**Include the full street address, city and zip code.**

| | Name of location | Street address |
|---|---|---|
| Location #1 | | |
| #2 | | |
| #3 | | |
| #4 | | |
| #5 | | |
| #6 | | |
| #7 | | |
| #8 | | |
| #9 | | |
| #10 | | |

**26. What types of Internet connections do the above facilities have? Please check all that apply.**

| | |
|---|---|
| DSL | |
| Cable modem | |
| T1 | |
| Fiber | |
| DS3 | |
| OC1, OC3, etc. | |
| Ethernet | |
| Satellite | |

| | |
|---|---|
| Wireless | |
| Other (please specify below) | |

In case of "other", please specify the type(s) of Internet connections used:

**27. What are the data rates of the Internet connections at each of these facilities? Please check all that apply.**

**In this case, we are interested in the total bandwidth for the connection between the facility and its Internet provider, not the download speed experienced by individuals at that institution.**

| | |
|---|---|
| 1 to 4.9 Mbps | |
| 5 to 49 Mbps | |
| 50 to 499 Mbps | |
| 500 Mbps to 4.9 Gbps | |
| 5 Gbps and above | |

**28. Which company(s) provide Internet service to the above facilities (public schools, community colleges/universities, public libraries and community centers)?**

| | |
|---|---|
| Provider #1 | |
| #2 | |
| #3 | |
| #4 | |
| Other(s) | |

**29. For those facilities which do have Internet connections, approximately how much did they pay to install the service, and what are the ongoing monthly charges?**

**Please give the complete range of prices which they are paying, for each range of speeds. For example, if one school is paying $100/month for a 10 Mbps connection, and another is paying $300/month, enter "$100 to $300".**

| | One-time installation / setup costs | Monthly cost |
|---|---|---|
| Below 5 Mbps | | |
| 5 to 49 Mbps | | |
| 50 to 499 Mbps | | |
| 500 Mbps to 4.9 Gbps | | |
| 5 Gbps and above | | |

**30. For those facilities which do have Internet connections, are they generally satisfied with the quality of their connections? If not, please indicate all reason(s) why.**

| | Public schools | Community colleges and universities | Public libraries | Community centers |
|---|---|---|---|---|
| Price is too high | | | | |
| Connection is too slow | | | | |

| | | | | |
|---|---|---|---|---|
| Service is unreliable | | | | |
| Not enough choice of providers | | | | |
| Other (please specify below) | | | | |

If there are other reasons for dissatisfaction, please enter them here:

**31. Other public access points. Please enter the addresses of any other public locations where people may access the Internet, aside from the places discussed above -- such as coffee shops.**

**Include the full street address, city and zip code.**

| | Name | Street address |
|---|---|---|
| Location #1 | | |
| #2 | | |
| #3 | | |
| #4 | | |
| #5 | | |
| #6 | | |
| #7 | | |
| #8 | | |
| #9 | | |
| #10 | | |

## 6.3. *Appendix – Technical Programming Notes*

This section contains a more detailed description of the procedure and programming tools we used to collect broadband availability data from the ISP websites. The data was collected during April and May 2008.

Both the Comcast and Verizon websites allow one to enter a street address and/or phone number in order to find out if broadband service is available. Using free off-the-shelf software tools, it is possible to submit a set of addresses to these websites in an automated fashion, and therefore do a rough automated survey of broadband availability. The software used is a kind of macro player running inside a web browser, which directs the web browser to visit a website, enters information in the form and "clicks" the button to submit the form. This is similar to the "macro" feature in Microsoft Word and Excel, which allows one to automate certain tasks. We used a free tool called *iMacros* by the iOpus Software company[56] which provides this macro facility for the Mozilla Firefox web browser.

After the browser submits the form, it waits for the response to come back from the server, and saves that result to an HTML file. It then returns to the webpage with the form, and repeats the process but enters different data into the form. This data can be read from a comma-delimited text file. Thus the final output of the iMacro script is a series of HTML pages, one for each of the records in the comma-delimited file. A further processing step is necessary to extract the relevant data from the HTML pages, to make it suitable for further processing. We used the freely-available Perl language[57] for this purpose, as it has very good facilities for processing text files.

In order to collect the addresses and phone numbers, we used a similar automated process to download data from a public White Pages directory website, whitepages.com. Ideally, one would like to download all of the listings (address and phone number) for a town, and then do a random sampling of that entire set, but the website does not make it easy to do that. It is not possible to search for all the entries in a town; instead, one must search for all the entries where the last name begins with a certain letter, such as S. Even then, it only displays the first 300 results, so for large towns, one will only see a limited portion of the entries, mostly starting with Sa, Sb, etc. Certainly there are some limitations with this approach as far as getting a truly random sample of households, but we felt it was still worthwhile for a demonstration project. With more programming effort, it would be possible to collect a more representative sample.

Some white pages listings do not include a phone number, which means that it is not possible to check broadband availability on the Verizon website: the website requires a Verizon landline phone number. Conversely, the Comcast website requires a street address. Some listings have P.O. Boxes, which therefore can not be checked on the

[56] http://www.iopus.com/imacros/ version 6.0.3.4

[57] ActivePerl http://activestate.com/Products/activeperl/index.mhtml for Windows XP

Comcast website, nor can they be mapped with GIS software. In addition, some listings on the whitepages.com site are for businesses, not residences, so these must be filtered out.

The complete process for "surveying" broadband availability by this method is as follows:

1. Download a set of HTML pages with address/phone number listings from the white pages website. (iMacros)

2. Extract the addresses/phone numbers from those HTML pages into a comma-delimited file. At this point, a unique ID is assigned to each location, to make it easier to track the location through the process. For example, an ID of "lexington-s5" would indicate the 5th entry in Lexington with a last name beginning with S. (Perl)

3. Submit the selected address/phone numbers to the provider websites and save the resulting pages. This produces a separate HTML page for each location. The name of the HTML file includes the unique ID for that location. (iMacros)

4. Read through the HTML pages to determine the broadband status of each location (available, not available, unknown, etc.) Generate another comma-delimited file with a field that includes the broadband status. (Perl)

5. Load the address file and the broadband status file into the GIS program. (ArcGIS)

6. Geocode the addresses, to display the dots with the status. (ArcGIS)

One may also load the comma-delimited files into Excel or other statistics software for data analysis.

One difficulty with this approach is that if any of the websites change their layout or user interface, it may be necessary to modify the scripts in order to work with the new look and feel. In other words, the scripts are brittle with respect to changes in the websites.

6.3.1. Source code and descriptions of each of these programs are in the sections below.

## iMacro scripts for checking broadband availability on provider websites

Here we give the macro code. The numbers at the side are not part of the code, they are simply used to make it easier to refer to parts of the code in our description.

```
1     VERSION BUILD=6030318 RECORDER=FX
2     TAB T=1
3     '
4     CMDLINE !DATASOURCE c:\temp\addrs.csv
5     SET !DATASOURCE_COLUMNS 11
6     SET !LOOP 2
7     SET !DATASOURCE_LINE {{!LOOP}}
8     '
9     SET !ERRORIGNORE YES
10    '
11    URL GOTO=http://www22.verizon.com/Residential/Broadband/
12    TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:HomePage ATTR=ID:txtAreaCode
      CONTENT={{!COL9}}
13    TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:HomePage ATTR=ID:txtPrefix
      CONTENT={{!COL10}}
14    TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:HomePage ATTR=ID:txtPhoneNumber
      CONTENT={{!COL11}}
15    TAG POS=1 TYPE=IMG ATTR=NAME:&lid=GO
16    WAIT SECONDS=6
17    SAVEAS TYPE=HTM FOLDER=c:\temp\providers FILE=verizon-{{!COL1}}.html
```

This code works as follows: lines 1 and 2 are simply generic initialization code. The lines beginning with an apostrophe ‘ are simply used to break up the code to make it more readable. Lines 4-7 describe how to read data from a data file. Line 9 says to ignore certain errors which would otherwise force iMacros to stop. Line 11 tells the web browser which web page to start from on the Verizon website. Lines 12 through 14 fill in the form with the values from the comma-delimited file.

The script for the Comcast website is similar:

```
VERSION BUILD=6030318 RECORDER=FX
TAB T=1
'
CMDLINE !DATASOURCE c:\temp\addrs.csv
SET !DATASOURCE_COLUMNS 12
SET !LOOP 2
SET !DATASOURCE_LINE {{!LOOP}}
'
SET !ERRORIGNORE YES
'
URL GOTO=http://www.comcast.com/default.html
TAG POS=1 TYPE=INPUT:TEXT FORM=ACTION:/localization/Localize.ashx
ATTR=ID:StreetName CONTENT={{!COL2}}
TAG POS=1 TYPE=INPUT:TEXT FORM=ACTION:/localization/Localize.ashx ATTR=ID:Zip
CONTENT={{!COL7}}
TAG POS=1 TYPE=INPUT:IMAGE FORM=ACTION:/localization/Localize.ashx
ATTR=ID:ShopProducts
WAIT SECONDS=15
SAVEAS TYPE=HTM FOLDER=c:\temp\providers FILE=comcast_{{!COL1}}.html
```

The script for downloading data from whitepages.com is as follows:

```
VERSION BUILD=6030318 RECORDER=FX
TAB T=1
'
CMDLINE !DATASOURCE c:\temp\whitepages.csv
SET !DATASOURCE_COLUMNS 2
SET !LOOP 5
SET !DATASOURCE_LINE {{!LOOP}}
'
SET !ERRORIGNORE YES
'
URL GOTO=http://www.whitepages.com/
TAG POS=1 TYPE=A ATTR=TXT:Advanced<SP>Search
TAG POS=1 TYPE=INPUT:CHECKBOX FORM=NAME:people_search ATTR=ID:name_begins_with
CONTENT=YES
TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:people_search ATTR=ID:name CONTENT={{!COL2}}
TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:people_search ATTR=ID:city_zip
CONTENT={{!COL1}}
TAG POS=1 TYPE=INPUT:TEXT FORM=NAME:people_search
ATTR=ID:people_search_state_id CONTENT=MA
TAG POS=1 TYPE=BUTTON ATTR=TXT:Search
TAB T=1
TAG POS=1 TYPE=A ATTR=TXT:view<SP>the<SP>first<SP>10<SP>results
TAG POS=1 TYPE=SPAN ATTR=TXT:Home<SP>*
TAG POS=1 TYPE=INPUT:RADIO FORM=NAME:filtering ATTR=ID:listing_category_home
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-1.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-2.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-3.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-4.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-5.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-6.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-7.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-8.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-9.html
TAG POS=1 TYPE=A ATTR=TXT:Next
SAVEAS TYPE=HTM FOLDER=c:\temp\mtcdata FILE={{!COL1}}-{{!COL2}}-10.html
TAG POS=1 TYPE=A ATTR=TXT:Next
```

### 6.3.2. Processing the white pages output

This script uses the freely-available *HTML::Parser* Perl module to process the pages downloaded from the whitepages.com website. The parser is configured to call a designated function whenever particular tags are encountered in the page. Based on the attributes of those tags, it is determined whether or not it is a tag that contains data which needs to be saved. When the end of a record is encountered, the data for that record is printed out, along with a unique ID. The comments in the code explain in further detail how the relevant tags are identified.

The following is an example of the output from this script. Some of the records are printed on multiple lines for readability; in reality, each record is contained in a single line.

```
id,street_address,extended_address,locality,region,postal_code,zip,phone_tel,ph
   onearea,phone1,phone2,
lexington-125,"36 Fairlawn Ln","","Lexington","MA","02420-2715","02420","(781)
   274-8735","781","274","8735",
lexington-126,"8 Hillside Ter","","Lexington","MA","02420-3405","02420","(781)
   538-5824","781","538","5824",
lexington-127,"18 Meriam St","","Lexington","MA","02420-3640","02420","(781)
   862-5923","781","862","5923",
lexington-128,"40 Ledgelawn Ave","","Lexington","MA","02420-3435","02420","(781)
   674-1186","781","674","1186",
lexington-129,"8 Fletcher Ave","","Lexington","MA","02420-3702","02420","",,,,
lexington-130,"4 Malt Ln","","Lexington","MA","02421-7031","02421","(781) 274-
   9022","781","274","9022",
lexington-131,"15 Allen St","","Lexington","MA","02421-7139","02421","(781)
   652-0312","781","652","0312",
lexington-132,"365 Woburn St","","Lexington","MA","02420-2306","02420","",,,,
lexington-133,"9 Sunny Knoll Ave","","Lexington","MA","02421-
   4341","02421","(781) 863-8618","781","863","8618",
lexington-134,"15 Pine Knoll Rd","","Lexington","MA","02420-1206","02420","(781)
   862-6575","781","862","6575",
```

```
#
# whitepages.pl ID-PREFIX START-NUM FILENAME ...
#
# Reads the HTML pages from the whitepages.com website,
# and produces a comma-delimited file containing street addresses and phone
numbers.
#

use strict;

use lib "../../lib/perl/lib/perl5/site_perl/5.8.3/";
use lib "../../lib/perl/lib/perl5/site_perl/5.8.3/sun4-solaris-thread-multi/";

use HTML::Parser;


my $idprefix = "";

my $recordid = -2000; ## should never get used
my $buffer = "";
my $saveit = 0;
my $curfield = "";
my %currecord = ( );

my @FIELD_CLASSES =
 ( "street-address", "extended-address", "locality", "region", "postal-code",
 "phone tel" );

my %FIELD_HASH = ( );
foreach my $f (@FIELD_CLASSES) {
 $FIELD_HASH{$f} = 1;
}


# textHandler
#
# this function is called when the parser finds normal text
# if we are inside an important tag, then we save the text we find
#
sub textHandler
{
 if ($saveit) {
 my $t = shift;

 $buffer .= $t;
 }
}

# tagHandler
#
# this function is called for start and end tags
# for end tags, the parameter will start with "/"
#
sub tagHandler
{
 my $tag = shift;
 my $attrhashRef = shift; # only for start tags

    if (($tag eq "span") || ($tag eq "p")) {
     # Beginning of a <span> or <p> block
    #

    # Look at the “class” attribute of the tag, to see if

     # it is one of the fields we are interested in
     #
 $curfield = $$attrhashRef{"class"};

 $saveit = (defined $curfield) && (defined $FIELD_HASH{$curfield});

    } elsif (($tag eq "/span") || ($tag eq "/p")) {
```

```
     # End of a <span> or <p>
     #
     # Save the text to $currecord if appropriate
     #
 $currecord{$curfield} = $buffer
 if ($saveit);

        $saveit = 0;
        $buffer = "";
 $curfield = "";


        } elsif ($tag eq "div") {
        # A <div> tag may mark the end of a record:
 # check the “id” attribute
 #
 # If it is the end of a record, print out the
 # values we saved
        #
 my $id = $$attrhashRef{"id"};

 if ((defined $id) && ($id =~ m/^nr_results_multiple_lisitng/)) {
 print $idprefix . $recordid . ",";

 foreach my $f (@FIELD_CLASSES) {
 my $val = $currecord{$f};
 print '"', (defined $val) ? $val : "", '",';

 if ($f eq "phone tel") {
 ## special treatment -- parse the phone number
 my $output = ",,,";

 if (defined $val) {
 if ($val =~ m/\((\d+)\)\s*(\d+)-(\d+)/) {
 $output = "\"$1\",\"$2\",\"$3\",";
 }
 }

 print $output;

 } elsif ($f eq "postal-code") {
 ## zip code -- get just the 5-digit ZIP
 if ((defined $val) && ($val =~ /(\d\d\d\d\d)/)) {
 print "\"$1\",";
 } else {
 print ",";
 }
 }
 }
 print "\n";

 %currecord = ( );
 $recordid++;
 }

 }
}



###########################################
#
# MAIN LOOP
#
#

if (@ARGV < 3) {
 die "whitepages.pl ID-PREFIX START-NUM FILENAME ...\n";
}

$idprefix = shift @ARGV;
```

```
$recordid = shift @ARGV;


# print the header line
#
print "id,";
foreach my $ff (@FIELD_CLASSES) {
 my $f = $ff;
 $f =~ s/ /_/g; # can't have spaces in column names
 $f =~ s/-/_/g;

 print $f, ",";

 if ($f eq "phone_tel") {
 print "phonearea,phone1,phone2,";
 } elsif ($f eq "postal_code") {
 print "zip,";
 }
}
print "\n";

# this parser will read through the input, calling the above
# functions for text and tags, and discarding everything else
#
my $p = HTML::Parser->new(
                     text_h => [ \&textHandler, "dtext" ],
                     start_h => [ \&tagHandler, "tag, attr" ],
                     end_h => [ \&tagHandler, "tag" ],
                     );

my $pattern;
while ($pattern = shift @ARGV)
{
 my $path;
 while ($path = glob($pattern)) {
       $p->parse_file($path);
 }
}

$p->eof();
```

### 6.3.3. Parsing output from provider websites

This script reads through a series of HTML pages download from the Comcast and Verizon websites, and looks for key strings of text which indicate the broadband status for a particular location. It produces a comma-delimited file containing the unique ID, town name, provider name and broadband status for each location. Here is an example of the output:

```
id,town,provider,availability
concord-l1,concord,verizon,not-available
concord-l10,concord,verizon,not-available
concord-l11,concord,verizon,unknown
concord-l12,concord,verizon,not-available
concord-l13,concord,verizon,RETRY
concord-l14,concord,verizon,available
concord-l15,concord,verizon,unknown
concord-l16,concord,verizon,available
concord-l17,concord,verizon,not-available
concord-l18,concord,verizon,available
concord-l19,concord,verizon,available
```

One difficulty with this approach to processing the HTML pages is that as soon as the providers change the design or language of this part of their websites, the script may no longer work. Also, it might be desirable to check the entire content of the files in some fashion, perhaps by comparing them against known and trusted samples pages, so that one can be sure there is no other information on the page which is also relevant.

```
# provider_output.pl
#
# Reads all of the Comcast or Verizon HTML pages in a directory,
# and checks for information about broadband availability.
#
 Outputs a comma-delimited file containing the broadband status
# found in each file.
#


print "id,town,provider,availability\n";



my $dirname;

while ($dirname = shift @ARGV)
{
 opendir(DIRH, "$dirname")
        or die "Couldn't open $dirname : $!";

 my $f;
 while ($f = readdir(DIRH)) {
 next unless ($f =~ /^(comcast|verizon)_?-?(.+)\.html/);
 my $provider = $1;
 my $fileid = $2;
 my $town = ""; ## first part of filename (before the dash) is the town
 if ($fileid =~ /^(.*?)-/) {
 $town = $1;
 }

 my $fullpath = ($dirname . "/" . $f);
 next unless (-f $fullpath);

 # Read the HTML file, line-by-line
 #
 # Look for certain “magic strings” which tell us the broadband status,
 # and make a note of each string we see.
 #
 open(INFILE, $fullpath)
 or die "Couldn't open $f : $!";

 my %flags = ( );

 while (<INFILE>)
 {
 # Verizon
 if ( /We\'re sorry\. Verizon Internet is not currently available/ ||
 /Verizon High Speed Internet for Business is not currently available/ ) {
 $flags{"not-available"} = 1;
 }

 if ( /We’re sorry, but we don’t recognize/ ) {
 $flags{"unrecognized"} = 1;
 }

 if ( /Sorry, we are unable to determine if Verizon High Speed Internet is
available/ ) {
 $flags{"unknown"} = 1;
 }

 if ( /Great News\! High Speed Internet is available\./ ||
 /Great News\!<br>Verizon FiOS Internet Service is available\./ ||
 /Good News\! High Speed Internet Is Available\./ ) {
 $flags{"available"} = 1;
 }

 if ( /Great News\! Verizon High Speed Internet is available for your business/
) {
```

```
 $flags{"available"} = 1;
 }

 ###### "Looks like Verizon ... is already on this line"
 if ( /like Verizon (High Speed|FiOS) Internet( Service)? is already on this
line/ ) {
 $flags{"already-haveit"} = 1;
 }

 if ( /Enter your home number to check if <BR>Verizon provides service in your
area/ ) {
 $flags{"RETRY"} = 1; ## didn't get past the front page
 }

 # Comcast
 if ( /Download music, photos and videos way faster than DSL and Dial up/ ) {
 $flags{"available"} = 1;
 }

 if ( /We are unable to find a match in our system for the address and ZIP Code
you entered\./ ) {
 $flags{"unrecognized"} = 1;
 }

 if ( /So we can provide you with the most accurate product availability and
pricing info possible, please enter your street address and ZIP code in the
fields below\./ ) {
 $flags{"RETRY"} = 1;
 }

 if ( /We're sorry, an error has occurred\./ ) {
 $flags{"RETRY"} = 1;
 }

 }

 close(INFILE);


 # Print the results
 #
 # check if more than one “magic string” was encountered,
 # or if none of them were encountered; in either case,
 # note an error
 #

 my @flaglist = keys(%flags);

 print "$fileid,$town,$provider,";

 if (@flaglist > 1) {
 print "\"ERR:" . join('/', sort(@flaglist)) . "\"";
 } elsif (@flaglist == 0) {
 print "NONE";
 } else {
 print $flaglist[0];
 }

 print “\n”;
 }

 closedir(DIRH);
}
```