

Notice and Request for Comments
U.S. Department of Commerce
National Telecommunications and Information Administration (NTIA)

Citizen Enabled Advances in Broadband Availability Data

Docket No. 180427421–8421–01

RIN 0660–XC042

Joint comments respectfully submitted by

The Quello Center
Michigan State University
404 Wilson Road, Room 406
East Lansing, Michigan 48824
Email: bauerj@msu.edu

and

Merit Network
1000 Oakbrook Drive, Suite 200
Ann Arbor, Michigan 48104-6794
Email: sawasky@merit.edu

Submitted via email to: mappingrfc@ntia.doc.gov

July 16, 2018

Citizen Enabled Advances in Broadband Availability Data

Broadband is of increasing significance to all consumers, government policy, economic development, scholarly research, community access, and education both inside and outside the classroom.¹ Measuring broadband availability is critical to its successful provision. This comment argues that current problems with the accuracy of broadband data can be resolved by implementing a consumer-sourcing mechanism for collecting individual level data. By combining and augmenting data sets, the weaknesses of any particular data set can be mitigated. Making these combined data sets open and publicly available will enhance their utility and reliability. The value and feasibility of this data collection will be enhanced by enlisting the help of regional and national networks that include educators, academics and information technology practitioners, advocates and consumers.

Considering that any source of data will have strengths and weaknesses, strategically using multiple sources of data can advance the quality of data to inform decision making. Specifically, data sources, such as FCC Form 477, can be analyzed in conjunction with new consumer-sourced data to improve the accuracy of broadband availability data and enable us to identify areas where access or speed appears to be under- or over-estimated. This approach is currently under development here in Michigan through a partnership between the Quello Center at Michigan State and the Merit Network. Our approach has the potential to be scaled up and sustained with the guidance and oversight of a National Advisory Group that will include federal agency representatives, industry, and current data source representatives.

The novelty of this recommendation is the combination of:

- Automatically-collected technical network and performance data combined with the simultaneous collection of consumer-provided demographic and local service information
- A single responsive web application for both fixed and mobile services to ensure coverage across all modalities
- Partnerships among a participating coalition of research and education networks that continually harvest and organize information facilitated by qualified researchers

Part I: The Problem

High-speed Internet connections are not equally available to everyone and the data to accurately measure the speed and reach of broadband is potentially misleading.² Estimates are particularly problematic in underserved areas such as rural and distressed urban areas of the United States, as

¹ Computer Science and Telecommunications Board, National Research Council (2002), *Broadband: Bringing Home the Bits* (Washington DC: National Academy Press); Broadband Opportunity Council (2015), *Report and Recommendations*. (Washington DC: NTIA, Department of Commerce); and FCC (2016), Federal Communications Commission, *2016 Broadband Progress Report* (Washington DC: Federal Communications Commission, 28 January); and Ofcom (2016), *Achieving Decent Broadband Connectivity for Everyone*. (London: Ofcom, 16 December).

² Jayakar, K., Maitland, C., Peha, J., Stover, S., and Bauer, J. (2016). *Broadband 2021, Report of the Interdisciplinary Workshop on the Development of a National Broadband Research Agenda*, July 25, 2016. Institute for Information. Available at https://broadband.ist.psu.edu/wp-content/uploads/sites/16/2016/09/Broadband-2021_Final_Report.pdf.

illustrated by research on Detroit.³ Current broadband data collection procedures and standards that often result in inaccurate results are making interventions and policy decisions more difficult.⁴ Right now data challenges that continually interfere with decision making include: the level of measurement, the over-reliance on a single source of data, and using ISPs as the major source of the data. Fortunately, challenges with inaccurate or incomplete information can be mitigated by complementing existing data sources (e.g., FCC Form 477, Broadband Catalysts, Measurement Lab and California's Public Utility Commission) with real-time data from consumers. Linking on-the-ground, consumer-sourced data with existing data will resolve key concerns about information accuracy while empowering broadband users in every region of the country.

Weaknesses with Current Approaches to Measurement

The Federal Communications Commission (FCC) and the National Telecommunications and Information Administration (NTIA) are the main sources of historical and current broadband availability data. However, the data being used to develop the FCC broadband map is based on a set of conventions that result in potentially misleading outcomes, such as overestimating the reach of broadband availability and speeds. Information is further stymied by limitations on data granularity and over-reliance on self-reports of Internet Service Providers (ISPs).

Granularity of Data

There are at least three key issues with current levels of measurement of broadband reach and speed. First, the current standard of measuring by census blocks is misleading. Since the FCC does not collect highly granular data, such as the locations of homes and businesses served, one subscriber in a census block presumes availability throughout⁵. However, service reported in a census block does not imply availability of service throughout the block, particularly for sparsely populated blocks with large land areas in rural regions. Second, with respect to the NTIA data, distinctions between residential and non-residential services are problematic, and therefore service (or service of a particular quality) may not actually be available to residential consumers in a census block where service is reported. Third, NTIA data reports speeds according to relatively wide tiers (e.g. 6-10 Mbps, 10-25 Mbps, etc.), which distorts reporting and effective categorization for speeds at the margins (e.g. approximately 10 Mbps)⁶. By bringing ground-level data generated by users to enrich the data collected by the FCC and NTIA, problems with granularity can be mitigated.

Self-reports

The FCC broadband map does not currently include on-the-ground measurements, thereby holding out the risk of over- or under-estimating broadband speeds and accessibility. More specifically, ISP data used to inform the broadband map are based on self-reported maximum advertised speeds by census

³ Reisdorf, B. C., Hampton, K., Fernandez, L., and Dutton, W. H. (2018), *Broadband to the Neighborhood: Digital Divides in Detroit*, Quello Center Working Paper. East Lansing, MI: Quello Center, MSU, January. Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3103457

⁴ Mack, E. A., Dutton, W. H., Rikard, R. V., and Yankelevich, A. (2018). *Mapping and Measuring the Information Society: A Social Science Perspective on the Opportunities, Problems and Prospects of Broadband Internet Data in the United States*, submitted to *The Information Society*, under review.

⁵ Bloch, S (2018, June 20). *Misleading data from internet service providers is cutting Iowa off from billions in broadband subsidies. Other states can be missing out, too. A New Food Economy investigation. The New Food Economy*, <https://newfoodeconomy.org/rural-iowa-broadband-data-fcc/>

⁶ Dutton, W. H., Reisdorf, B., Rikard, R. V., Shapiro, M., and Yankelevich, A. (20??). *A New Social Science Perspective on Digital Inequalities: Embedding Policy, Infrastructure, and Place*.

blocks. However, advertised speeds are not necessarily representative of available or average speeds faced by consumers. The threat to inaccurate measures are particularly problematic for a “distance sensitive” technology such as DSL, which varies based on distance and line quality. In fact, these self-reported estimates do not match the speeds independently measured by third party agencies like the Open Technology Institute⁷. Issues resulting from self-reported data from ISPs can be mitigated by including measures based on reports from individual users, applying crowd-sourcing techniques, as outlined later in this comment.

Missing Data and Quality Management

Accurate measures of broadband availability are often compromised because data are extrapolated from sources that are not specifically designed for the purpose of measuring broadband availability. Also, there is no requirement to explain how these data are collected, making it difficult to assess their reliability. More specifically, in their filings related to FCC Form 477, publicly subsidized providers are not required to disclose their data collection process or how services are advertised in a specific market.

Beyond quality management, missing data also impacts communities, as multiple examples in rural and distressed areas illustrate. For instance, the FCC’s national broadband map indicates that every county in Iowa is covered by download speeds of 25 megabits per second (Mbps). However, when these data are combined with and compared to speed tests facilitated by the Open Technology Institute a different finding emerges — residents experience speeds of 25 Mbps only 22 percent of the time. This illustrates how current practices can yield inaccurate data that can be misleading for policy and regulation.

While some academic studies have sought out alternative sources of data from the private sector that fill information gaps in public databases,⁸ there are methodological issues with using private data sources, particularly if there are information asymmetries about how data were collected and processed⁹. For example, there are inherent incentives for some private providers to over- or underestimate the provision of broadband. This problem would be mitigated with an agreed upon validation process and an adaptable and flexible data collection design as outlined in a later section of this comment.

The Exclusive Nature of Broadband Data

The over-reliance on ISP’s as the major source of data to map reach and speed is problematic for several reasons. While there are wide differences between ISPs, they are typically not transparent about their practices and not required to submit all known data points or commercial interests. Another problem with this reliance is that it distances communities from the data collection and diminishes the data accessible to local decision makers. This distance inevitably muddies decision making at the community level. This distance prevents local decision makers and advocates from making rational, data-driven decisions about the best interests of their communities.

Alternatively, with advances in big data and computational analytics, it is feasible to open up the data collection to include individual consumers. By enabling people at the local level to take on a more

⁷ Bloch, S (2018, June 20). Misleading data from internet service providers is cutting Iowa off from billions in broadband subsidies. Other states can be missing out, too. A New Food Economy investigation. *The New Food Economy*, <https://newfoodeconomy.org/rural-iowa-broadband-data-fcc/>

⁸ Grubestic, T. H. (2015). The broadband provision tensor. *Growth and Change*, 46(1), 58-80.

⁹ Malone, C, Nguyen, M. (2018). We Used Broadband Data We Shouldn’t Have – Here’s What Went Wrong. *FiveThirtyEight*. Available at https://fivethirtyeight.com/features/we-used-broadband-data-we-shouldnt-have-heres-what-went-wrong/?mc_cid=2680ac13ec&mc_eid=558f45e107.

meaningful role in shaping outcomes concerning broadband availability and quality, local stakeholders can be empowered to take ownership of the future of connectivity in their neighborhoods. In addition, by enlisting the help of the public through ground-level broadband data collection, routine data collection can spawn learning and greater interest of issues related to broadband access and quality. Relatedly, by making the data more publicly available, we expect to see an increased level of confidence for rational decision making and greater investments with reduced risks. In other words, when people are removed from the data collection process and do not understand or trust the quality and accuracy of the information given to them, investments are reduced and progress is delayed.

Successful Crowd Sourcing and Illustrative Examples from Business and Citizen Science

The potential of crowd sourcing has been well developed by several major scholars and practitioners.¹⁰ As is the case with any data collection method, crowd-sourcing data brings with it a number of challenges that need to be considered and controlled through strong management of data collection and analysis. For example, one main challenge will be that highly skilled Internet users, who have better connections, will be overrepresented. Therefore, unweighted or overly simplistic aggregation of all data can yield over or under representations of broadband availability and speeds. Because of these types of challenges, accuracy of the data requires validation methods devised and checked by reputable researchers.

Crowd sourcing is somewhat of a misnomer, in that successful uses of this approach are well managed 'collaborative networked organizations', or CNOs.¹¹ More generally, collaborative network organizations are ideal for aggregating geographically distributed information and intelligence, and supporting greater independence of, and less control by, established institutions. While crowd sourcing is a popular expression, it is not an outcome of uncontrolled or disorderly crowd behavior. To the contrary, successful uses of crowd sourcing are well managed. There are CNOs focused on simply sharing data among participants, such as InnoCentive, Bugzilla and Neurocommons; others are focused on obtaining contributions from many participants, such as Zooniverse¹², which asks people to classify galaxies, and news aggregation sites, such as Digg, medical information sites, such as Sermo, and information markets; others are devoted to co-creating a product, such as Wikipedia, and Firefox, open sourced software developed through collaboration.

The FCC's work on using a mobile app to measure broadband performance is perhaps the most relevant US example. It has been successful and can easily be mined to complement the proposed data set. The proposed data collection in this comment will leverage the success of the FCC and the network of organizations involved with Merit (Michigan's Research and Education Network, see Appendix A) inside Michigan initially - and other participating national Research and Education Networks post-pilot to

¹⁰ See, for example, J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, Doubleday, New York, 2004; Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven, CT, 2006; S. E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press, Princeton, NJ, 2007; and C. R. Sunstein, *Infotopia: How Many Minds Produce Knowledge*, Oxford University Press, New York, 2006.

¹¹ W. H. Dutton, W. H. (2007), 'Through the Network of Networks - The Fifth Estate', Inaugural Lecture, Examination Halls, University of Oxford, Oxford Internet Institute, Oxford, October 15: see <http://ssrn.com/abstract=1134502>.

¹² <https://www.zooniverse.org/>

manage data collection across the US in ways that can not be replicated easily by the FCC. Both within Michigan's pilot and if scaled to the national level, this proposed CNO will benefit from incorporating a broad set of data sourcing, management partners and representatives of all major stakeholders in a National Advisory Group. At the same time, potential areas where crowd-sourcing is above or below average (or even non-existent) can be identified by the use of multiple data sets and specifically targeted by researchers to mitigate these issues. Researchers can conduct detailed data gathering in specific areas identified as problematic.

Part II: A Proposed Solution - A New Network of Networks Approach to Collaborative Data Collection

We recommend a new approach for data collection that builds on collaborative network organizations (CNOs), such as approaches to citizen science¹³, to uniquely leverage (1) networks of stakeholders (i.e., Merit and other participating Research and Education Networks) to manage the sourcing of data from users across the nation; (2) a partnership with academic researchers at the MSU Quello Center that allows for data quality control (identifying and correcting problematic data) and sophisticated analyses using multiple sources and forms of data; (3) including data collected through an app designed to be flexible and adaptable to allow for multi-purpose analysis and use in the short and long term through any type of device, fixed or mobile. This approach will be scaled-up to the national level by creating a National Advisory Group to oversee mapping efforts that includes federal agency representatives, industry and current data source representatives.

As noted above, successful approaches to crowd-sourcing are not actually based on 'crowds' per se, but on complementary contributions by well managed CNOs, such as those that have produced Wikipedia.¹⁴ The FCC itself is successfully deploying an FCC Speed Test app for Android and iPhone devices to crowd source data on mobile broadband performance.¹⁵ While the potential of such an approach has been demonstrated, it is inherently limited by the FCC's physical location in Washington DC, and the potential for stakeholders to question its independence.

Is such an approach feasible?

Part III: Feasibility

The partners submitting this comment believe that a collaborative network organization can be effectively employed to enhance the quality of broadband data on accessibility in ways that will enhance learning and engagement by consumers. This is possible through a novel approach to this networking effort.

In the first instance, Merit Network — an educational research network based in Ann Arbor, Michigan¹⁶ — is working with the Quello Center at Michigan State University¹⁷ (MSU) to demonstrate the viability and value of our approach to consumer sourcing of broadband data. Working with MSU's Quello Center, Merit's collection of ground-level data collected across the state can be managed by an independent

¹³ Citizen science includes a variety of approaches to involving the public in research and has been the focus of major advances in research methods (Dutton, W. H., and Jeffreys, P. (2010) (eds), *World Wide Research: Reshaping the Sciences and Humanities*. Cambridge, MA: The MIT Press; and Meyer, E. T., and Schroeder, R. (2015), *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, MA: MIT Press).

¹⁴ Dutton, W. H. (2008), 'The Wisdom of Collaborative Network Organizations: Capturing the Value of Networked Individuals', *Prometheus*, 26(3), September, pp. 211-30.

¹⁵ <https://www.fcc.gov/general/measuring-broadband-america>

¹⁶ <https://www.merit.edu/>

¹⁷ <http://quello.msu.edu/>

research institution to demonstrate the ways in which data can be collected, cleaned (corrected) and used in meaningful analyses. Once the value of this approach is demonstrated through a pilot, the data collection can be moved statewide through Merit’s network, and on the basis of the success of this statewide rollout, it can then be scaled to the national level through enrolling other Research and Education Networks¹⁸ across the nation —with leadership and support from Merit and the Quello Center. Other Research and Education Networks can join as partners, contributors, and advisors to form the core of the envisioned CNO.

To enable this data collection, our collaboration is already in the process of developing a web-based software application and open analytical repository to be designed, developed, and deployed on a pilot basis. As novel and key elements of this proposal, the web application will be used by consumers and households to “crowd-source” data collection. In addition to a standard automatic “speed test” and related low-level technical data acquisition (e.g., geolocation, IP address owner, date/time stamp, browser used, device type, etc.), this application will also allow simultaneous end-user-supplied data collection, such as a qualitative rating of the service, select demographic information, and information about general service options in the locale.

The Design and Deployment of an Open Analytical Repository

The software will be designed as an HTML5, responsive web application, allowing intuitive use on all devices: computers, laptops, tablets and mobile phones. The collected data will be stored in an open and extensible database format, allowing qualified researchers to integrate other important databases for analysis and reporting. It is important that qualified researchers guide the initial crowd-sourced deployment, such as by creating a network of volunteers through agricultural extension officers, 4-H clubs, local libraries and college or high school students.

Collaborative Design, Collection and Analysis

This multi-method, multi-trait approach adds features for adaptability and accuracy with the unique strategic advantage of networked industry and practitioner partnerships that include an academic unit that understands approaches to citizen science as applied to practical problems. Partners in this effort include Merit, the Quello Center at Michigan State University, and potentially participating Research and Education Networks through The Quilt. Merit is a non-profit, member-owned organization governed by Michigan’s public universities, which owns and operates America’s longest-running regional research and education network. The Quilt is the national coalition of non-profit U.S. regional research and education networks representing 40 networks across the country. The James and Mary Quello Center has a 20-year history of conducting policy and media research, housed in the College of Communication Arts & Sciences, Department of Media and Information at Michigan State University. See Appendix A.

As noted above, if this work is deployed at the state or national scale, we propose forming a State or National Advisory Group, respectively, to support this effort. Merit and the National Advisory Group will help design and deploy the web application, and allow for revisions and further research. As lessons are learned, the application can be rolled out on a wider scale to other Internet users among the general public. Data quality management will involve faculty from multiple departments, anchored in the Quello Center, Michigan State University, where the core research staff and students will be co-located. The Quello Center can help oversee collection and cleaning of data to ensure its accuracy and also the

¹⁸ <https://www.thequilt.net/about-us/>

validity of any analyses of these data, which would be made available to wider research and practitioner communities.

Educational Outreach and Economic Development

A spin off of this activity will be greater community or public involvement in gauging broadband access in ways that can enhance learning and reshape attitudes and beliefs about broadband in positive ways. Technical infrastructures are not the only barrier to the up-take of broadband.¹⁹ The cultures, or attitudes and beliefs, of users are also a challenge. Research conducted by the Quello Center and others has found that if individuals are uninterested in getting online or do not see a value in being online, they are unlikely to overcome other barriers, such as acquiring the skills or other resources to take that step. Employing a web application to check on the availability and quality of broadband services will help individuals understand these issues, and help develop skills of value in many other areas, from health monitoring to environmental sensing, to issues related to the “homework gap”.

Expected Benefits

Existing data sources can be combined with the proposed applications new database for analyzing managed ground-up crowd-sourced data. Other relevant data sets include: the national broadband availability dataset (also referred to as the State Broadband Initiative (SBI) or National Broadband Map (NBM) dataset), the Census Bureau’s American Community Survey (ACS) and The Bureau of Labor Statistics. By integrating large national and regional datasets on broadband infrastructures with data on at-home broadband speeds and household level accessibility, we can advance the ability of policymakers and academics to empirically analyze numerous issues revolving around the Internet.

Among other things, the integration of these multiple sources will allow policy makers and academics to:

- Reduce the financial risks for community economic development decisions with clear indicators of unconnected and otherwise marginalized areas at the local level.
- Preserve resources by concentrating field research in areas that exhibit atypical or missing data points.
- Improve rationality for business decisions, for example by assessing the impact of competition on Internet quality.
- Track change across indicators over time at more granular levels of analysis to visualize change in relation to policy and regulatory initiatives, such as stimulus funding.
- Track the effect of new policies on Internet accessibility (digital divides) and quality (digital inequalities), while also accounting for geographic and individualistic characteristics.
- Move beyond taxonomy and characterization of various government initiatives²⁰ and on to empirically grounded policy research²¹.

¹⁹ Dutton, W. H., and Blank, Grant (2015), ‘Cultures on the Internet’, *InterMedia*, Winter 2014/15 Vol 42 Issue 4/5: 55-57; Reisdorf, B. C., Blank, G., and Dutton, W. H. (in progress), ‘Internet Cultures and Digital Inequalities’, pp. forthcoming in Graham, M., and Dutton, W. H. (eds), *Society and the Internet: How Networks of Information and Communication are Changing our Lives*, 2nd Edition. Oxford: Oxford University Press.

²⁰ Hauge, J. A., and Prieger, J. E. (2010). Demand-side programs to stimulate adoption of broadband: What works? *Review of Network Economics*, 9(3), 65-79.

²¹ Picot, A., and Wernick, C. (2007). The role of government in broadband access. *Telecommunications Policy*, 31(10), 660-674.

- Move beyond an overly simplistic urban/rural divide to compare well-to-do and distressed areas of rural and urban areas and other gradual differences across geographies.

Proposal Highlights

In conclusion, we are embarking on the development of a well-designed and well managed collaborative networked organization (CNO) to consumer-source data on broadband availability that could serve as a model and a framework for a systematic, nation-wide effort. This new source of data will help mitigate the problems with granularity, missing data, and the accuracy of currently available data. This approach is a policy-oriented application of citizen science approaches that have been successfully used by the FCC and others in comparable ways. By augmenting FCC, NTIA, and other publicly available data with ground-level data generated by users, problems with granularity can be avoided. The merging of data sets and sources provides a more meaningful analysis and more informed decision making.

More meaningful and context-driven data is possible with an adaptable and flexible collection design as outlined in this proposal. The technical aspects of designing and managing an open analytical repository, will be addressed by technicians and experts within the Merit Network and Quello Center. The success of this prototype test case will enable further nationwide deployment, the design of best practices and research opportunities facilitated by participating research and education networks, Merit and MSU's Quello Center. At both the state and the national level, an Advisory Group can facilitate broader participation by leading and overseeing efforts. The commenters welcome the opportunity to explain their approach to the NTIA and other interested stakeholders and possibly secure additional partners and support.

Appendix A: Partners

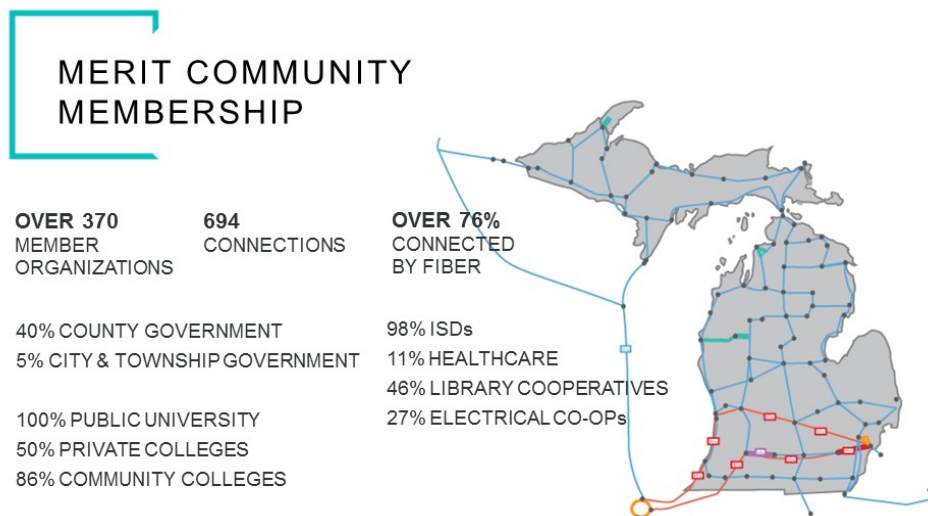
Merit Network, Inc. (Merit)

Merit Network, Inc. is a non-profit, Member-owned organization governed by Michigan's public universities. Founded in 1966, Merit owns and operates America's longest-running regional research and education network. After 52 years of innovation, Merit continues to provide high-performance networking and services to community anchor organizations including higher education, K-12, library, health care, government, research, and other nonprofit and service organizations.

Merit manages the nation's longest-running private, regional network. Beyond its core multiple 10Gbps Ethernet backbone, Merit maintains metropolitan fiber rings in Detroit, Grand Rapids, Lansing and Chicago. Beyond providing cost-effective bandwidth, a facilities-based network allows greater engineering control for Michigan institutions looking to exchange traffic over a private network. Merit is also Michigan's GigaPoP operator, connecting Michigan universities to the Internet2 network. Merit also provides sponsored Internet2 access to Michigan government organizations, K-12 schools and libraries so students and government employees across the state may benefit from the programs offered.

Through Merit, organizations have access to leading-edge network research, state and national collaborative initiatives and international peering. Merit is a trusted source for providing high quality network infrastructure, initiating and facilitating collaboration, and providing knowledge and technology transfer through outreach. Pursuit of this goal benefits a wide audience in the education, government and healthcare communities and has broad implications for economic development in Michigan.

Merit's mission is to be a respected leader in developing and providing advanced networking services to the research and education community. Merit believes in the strength of a robust community - By connecting organizations, building relationships and sharing services, Member institutions can achieve far more together than they can alone. Merit is headquartered in Ann Arbor, Michigan. More information at: <https://www.merit.edu/>



MSU Quello Center

The Quello Center seeks to stimulate and inform debate on the economic and social implications of media, communication, and information innovations of our digital age. It conducts research that questions taken-for-granted assumptions and collaborates with other centers of excellence in research on Internet studies, new media, and the social and economic implications of communication, media, and information technologies of our digital age as well as the policy and management issues raised by these developments. The Advisory Board includes dozens of high-profile industry and public sector thought leaders. As such, it is positioned to call upon its network of researchers, fellows, and board members to develop best practices, clean data, ensure the accuracy of data, and conduct meaningful analyses targeted toward informing and shaping policy and practice. More information at:

<http://quello.msu.edu/>