Challenges in measuring the Internet for the public Interest

David Clark and kc claffy
ddc@csail.mit.edu and kc@caida.org

September 2021

DRAFT: Working version. Please contact authors for latest version.

Abstract

The goal of this paper is to offer framing for conversations about the role of measurement in informing public policy about the Internet, the barriers to gathering measurements, public policy challenges that are creating pressure for reform in this space, and recommended actions that could facilitate gathering of measurements to support policy-making.

Contents

1 Motivation

Society is at a turning point with respect to the Internet. Since the mid-1990's, the private sector has led development of the Internet. Now, as the Internet has become critical infrastructure, public policy concerns are becoming more visible and important. The Internet today exposes users to a range of harms, including those arising from architectural limitations, poor security practices, performance impairment, consolidating industry structure, and the digital divide, that are serious enough to create public interest in mitigating them. Such mitigation is best initiated with a thorough understanding of the harms, including their scope, extent, and operational contexts in which they arise. The rising influence of adversarial actors is increasing the urgency for a more rigorous understanding of the Internet than today's ecosystem allows.

Government oversight plays an important role in other areas of critical concern to society, such as health care, transportation, ocean and atmosphere, food and drugs, and traditional telecommunications. An important part of this oversight is gathering data to understand how each system is working. But no actor today provides such oversight or data-gathering function for the Internet. As a result, operators, policy makers, scientists, and citizens today have no consensus view of the Internet to drive decision-making, or understand the implications of current or new policies or technologies. The quality and quantity of independent research on Internet infrastructure is deeply impaired by a lack of access to relevant data. This situation will get worse, and the lack of rigorous scientific research on the character of the Internet will grow even more problematic, as the Internet continues to be ever more deeply embedded in our lives. The goal of this paper is to offer framing for conversations about the role of measurement in informing public policy about the Internet, the barriers to gathering measurements, public policy challenges that are creating pressure for reform in this space, and recommended actions that could facilitate gathering of measurements to support policy-making.

1.1 The Measurement Challenge

It might seem that one could understand the Internet's operational character from analysis of its specifications. This is not so. The Internet is a complex system, its behavior an emergent consequence of operational decisions of thousands of independently operating Internet service providers, tens of thousands of edge-connected service providers and developers, and billions of users. Understanding the character of the Internet requires measurement of its behavior.

Moreover, in most of the world, the Internet infrastructure is the product of the private sector. Economic considerations that drive the private sector shape the character of the Internet, key aspects of its resilience, security, privacy, and its overall future trajectory. These considerations do not generally include support for independent scientific study of the infrastructure. Independent third parties can measure the Internet from the edge, draw their own conclusions, subject these to comparison and peer review, and publish results. But often edge measurements allow for only *inference* of properties or behavior, but not direct assessments.

Network operators collect substantial data on their own networks, but typically with a narrow focus and almost always limited availability and corporate interest in the messaging, which leads to concerns about bias in the reported data. Since the data may reveal aspects of business practices that operators hold close, they have strong counter-incentives to provide any data in the first place. Both factors have had troubling consequences for the research community. Sometimes a group of researchers manages to negotiate a one-time data-sharing agreement with a commercial firm, and get access to such data in order to perform research. Some papers resulting from this sort of collaboration report important findings. But often an employee of the associated company is an author on the paper, and the data is not available to other researchers for replication of results, triggering concerns regarding scientific objectivity.

Governments could gather data directly, but the trans-national character of the Internet raises challenges for government coordination. An accepted approach to data gathering and analysis is to make sure that data is made available to neutral third-parties such as academic researchers, who can independently pursue their efforts, draw their own conclusions, subject these to comparison and peer review, and present results as advice to governments.

Another option is mandated data reporting. Compelled reporting of data, and its analysis, comes at considerable cost to all parties. Such efforts thus requires much stronger justification than satisfying scientific exploration. In fact, government-mandated data is often encumbered, its use limited to purposes established in regulation, and thus not available for scientific use. This difference in activation thresholds between regulatory oversight and scientific inquiry poses a formidable challenge, because the 25-year gap in regulatory oversight has led to calls for objective scientific input to public policy debates. This is typically a role for (at

least) academic researchers.

Regardless of whether the institutions collecting and analyzing data are private sector, public sector, domestic, international, or some combination of these, the execution must effectively address a variety of technology, privacy, and business issues. Technical issues include speed, scale, and the distributed character of the Internet. Legal and ethical include navigating legislation governing data privacy, which require legal and technical expertise to interpret.

Finally, no governmental organization has sustaining Internet measurement as part of its mission. The FCC plays a central role in compelling providers to report data about service availability as part of its regulatory authority to promote broadband deployment,[1] which we discuss further in Section 4.1. But to gather measurements for scientific study, academics must obtain funding for data collection from organizations such as the NSF, which have research as their mission, not data collection. Advances in measurement in the public interest will have to address these challenges: objectivity of measurements and associated inferences; legitimate business interests in secrecy; respect for privacy, the role of the research community, and sustainability. These considerations will shape realistic future options.

1.2 Historical Context

Few people outside of Internet historians know that the first Internet backbones were created to connect scientific researchers to high performance computing facilities, and that the first general purpose Internet backbone was funded by the National Science Foundation (NSF) in the 1980s and 1990s. The NSFNET backbone fostered the intermediate evolution of the TCP/IP protocols, as it allowed an operational network to scale to millions of users. One of the authors (Clark) was one of the designers of the Internet protocols during this time. The NSF also supported one of the first scientific studies of this backbone,[2] relying on traffic, topology, and performance data that the NSF mandated be collected and shared publicly. The year after Claffy *et al.* published this study, the NSF decommissioned this backbone. The U.S. government launched complementary ambitious industrial policies to promote competition, and thus innovation, in the emerging Internet transport and domain name industries. The policy goal was to transition the Internet operations industry to the private sector, and make it a commercial undertaking.

This transition left the world with an Internet architecture not prepared for all the malicious actors that would try to exploit its weaknesses. First, its academic and scientific research roots led to a protocol architecture that assumed adversaries would not be operating parts of the infrastructure itself, and thus the protocols did not require authentication of addresses, routes, and names. Once it was clear how universal the Internet infrastructure would become, the Internet engineering standards community proposed technical solutions to retrofit layers of authentication into these protocols, but the solutions have not overcome the misaligned incentives, rooted in cost and complexity, that hinder deployment. It is easy to understand why profit-seeking firms may not be able to justify investment to enhance security. But that realization does not yield a clear path forward.

Second, securing fundamental mechanisms of the Internet requires some level of global governance to guarantee consistent interpretation of addresses and names. As part of the commercial transition, and "lessening the burdens of government",[3] the U.S. government led the private sector in establishing ICANN as the private, multistakeholder organization responsible for global coordination of the Internet identifier systems for the infrastructure industry, including preserving their security and stability.[4] Similar competitive pressures that inhibit investment in security have challenged this multistakeholder model of governance of Internet identifiers.[5]

[1] U.S. Congress. *Telecommunications Act of 1996, Pub. L. No. 104-104, 706(a), 110 Stat. 56, 153 (codified as 47 U.S.C. 1302).* 1996.

[2] claffy, k., G. Polyzos, and H. Braun. "Application of sampling methodologies to wide-area network traffic characterization". In: *ACM SIGCOMM*. May 1993.

[3] Internet Corporation for Assigned Names and Numbers. *ICANN Articles of Incorporation.* `https://www.icann.org/resources/pages/governance/articles-en`.

[4] Internet Corporation for Assigned Names and Numbers. *ICANN Bylaws.* `https://www.icann.org/resources/pages/governance/bylaws-en/`.

[5] Brian Cute. *A conversation about evolving the effectiveness of our multistakeholdermodel.* `https://www.icann.org/en/system/files/files/draft-evolving-multistakeholder-model-issues-list-25apr19-en.pdf`. Mar. 2019; Internet Corporation for Assigned Names and Numbers. *Board Action on Competition, Consumer Trust, and Consumer Choice Review.* `https://www.icann.org/news/blog/board-action-on-competition-consumer-trust-and-consumer-choice- review`. Mar. 2019; Internet Corporation for Assigned Names and Numbers. *Board Action on Security, Stability, and Resilience 2 Review.* `https://www.icann.org/en/system/files/bm/rationale-ssr2-22jul21-en.pdf`. July 2021.

2 A review of harms and remedies

In prior work,[6] we undertook a classification of Internet-related harms, by layer or segment of the ecosystem in which they arise. We took a structural approach to that analysis—where in the ecosystem does the harm arise, and what actors bear responsibility for the harm, or are best positioned to mitigate the harm. We defined a harm as *an impairment – either with respect to an individual, a firm or society – to an entity's welfare interests, relative to the normal expectations of the time and context.*[7] This definition reminds us that what constitutes a harm will evolve over time, such as what bandwidth is sufficient to consider a home broadband-connected. Our goal was a framework to help illuminate the interactions and tradeoffs – conflicting articulations of welfare interests – in attempting to mitigate any specific harm.

We also considered how unique harms on the Internet are to the Internet ecosystem. Some harms have a corresponding harm that pre-dates the Internet; the new question is how to define and detect its occurrence in a digital ecosystem, e.g., fraud or forgery. Other harms are well-known, but the accelerating and amplifying power of the Internet as a platform triggers debate as to how regulatory intervention might mitigate the harm. Targeted advertising, surveillance, and risk of addiction to video games or television have been around much longer than the Internet, but when the scope, precision, and potency expand by orders of magnitude, the potential for new harm emerges. Other harms are unique to a digital environment and its political economy, such as attacks on networks and digital information.[8] In this section we summarize the classes of harms we taxonomized in that paper, to frame our subsequent discussion of barriers to measuring them (Section 3).

2.1 Harms related to Internet access

The U.S. government's National Broadband Plan in 2010 implied that exclusion from broadband service is a harm: "All Americans should have access to broadband service with sufficient capabilities, all should be able to afford broadband and all should have the opportunity to develop digital literacy skills to take advantage of broadband.".[9] Dimensions of this harm include:

- **Lack of universal access.** To those without access, the now undeniable harm is marginalization with respect to the growing fraction of civil, commercial, and social activity that occurs online, sometimes primarily online. The measurement problem relates to accuracy of mapping. The general form of the remedy involves direct involvement and investment by the public sector to complement what the private sector can justify.

- **Low-quality service.** Independent measurements can gather some aspects of service quality. Competition in access service might stimulate providers to improve their service, but low service quality may occur where competition is not thriving. Where the public sector invests, specification of required service quality can prevent this harm. If the harm arises in the context of a privately-built network, the only remedy may be direct public sector investment to augment infrastructure capabilities.

- **High-cost ("unaffordable") service.** "Affordable" is a tricky term – it is not obvious what price sufficiently mitigates the contribution of price to the harm. If high prices result from lack of competition, the remedy is either price regulation (if the cost is unjustifiably high) or direct public sector investment to cover costs, which may be the correct approach in rural areas. Tracking service prices is necessary to understand where and in what context high prices raise an issue of public interest.

- **Low adoption.** Governments can mitigate this harm through education, and subsidies where cost is the barrier to adoption. The population of non-users may also naturally shrink over time.

- **Insufficient resilience.** This harm is more complex. There are many aspects of resilience, and different views about which metrics are most important. Measurement is difficult due to the distributed

[6] D. Clark and k. claffy. "Toward a Theory of Harms in the Internet Ecosystem". In: *Telecommunications Policy Research Conference (TPRC)*. Sept. 2019.

[7] We took this definition of harm from Kleining (Kleinig. "Crime and the Concept of Harm". In: *American Philosophical Quarterly* 15.1 [1978]).

[8] Anderson, Ross and Barton, Chris, and Bohme, Rainer, and Clayton, Richard, and van Eeten, Michel and Levi, Michael, and Moore, Tyler and Savage, Stefan. "Measuring the Cost of Cybercrime". In: *Workshop on Economics and Information Science*. `https://www.econinfosec.org/archive/weis2012/papers/Anderson_WEIS2012.pdf`. 2012.

[9] Federal Communications Commission. *The National Broadband Plan: Connecting America*. http://download.broadband.gov/plan/national-broadband-plan.pdf. 2010.

and decentralized nature of service provision in the ecosystem, although much consolidation has happened over the last two decades. Still, each service provider makes its own decisions about degree of redundancy, degree of interconnection with other regions, etc. Section 4.2 focuses on this complex and multi-dimensional metric.

2.2 Harms related to security

These harms manifest as incidents that affect specific users, but they also cause a general loss of trust in the system, which degrades its utility and value.

- Physical and network layer. Three critical systems in the Internet lead to misrouted and misdelivered traffic: BGP, the DNS and the CA system. End nodes have only a limited ability to protect themselves from these harms. The state of all three of these systems can in principle be measured, but with great effort, probably requiring a dedicated organization, and cooperation (voluntary or mandated) of many actors in the ecosystem. The decentralized character and lack of coordination hinders the viability of many proposed remedies to abuses of these systems. In the case of the CA system, the CA/Browser Forum [10] tracks the behavior of the various CAs and declares those that misbehave untrustworthy. Google has used its position in the market to create the Certificate Transparency system, which requires that when a CA issues a certificate, it posts a copy in a log that is public. This does not prevent a CA from issuing a false certificate, but it makes the event public, so a misbehaving CA cannot do so in secret.

- Edge devices and routers. The penetration of an edge device can lead to many consequential harms, including ransomware and other malware, which we classify as harms at the application layer. Tracking of malware and system vulnerability is a well-developed current practice. However, identification and measurement of harm is difficult. A pragmatic approach is to classify a system penetration as a harm, independent of the consequence of that penetration—an example of a *proxy harm*.

- Application layer. We identified three baskets of harms at the application layer. The first is malicious applications, perhaps downloaded from an app store, which appear to offer some useful service but which also cause harms, such as theft of personal information. Malicious applications present obstacles to detection. It is often hard to tell what an application is doing, especially if communication is encrypted. The second arises if providers of dominant applications add unwelcome behaviors, terms or conditions to the application. Such behaviors (if documented) would not qualify as malicious, but may still cause harm. If an application is dominant, and users persist with use of the application, the potential for this harm will increase. Examples include rent-seeking if the application is fee-based, or increasingly intrusive tracking of users and precision targeted advertising. The third category is applications that, while not themselves malicious, allow interactions among users without adequate protection against malicious users. There is no general method for assessment of such harms, in part because there is no clear specification of what constitutes a harm.

  The primary responsibility for mitigating these harms has thus far fallen on application designers, who must proceed pragmatically as behaviors emerge that are deemed harmful by governments or advocacy groups. Industry groups may attempt to self-regulate these harms through codes-of-conduct. Government agencies, e.g., FTC, may intervene as necessary. There will always be an underlying definitional tension as to what constitutes a harm, and what sort of data serves as legitimate evidence that the harm is occurring.

2.3 Harms to confidentiality (individual and organizational)

Privacy is an abstract concept in today's personal-data-driven ecosystem, and loss of privacy is not generally amenable to measurement. Law and regulation have provided proxy harms, which include violation of requirements for notice and consent, and data breach, i.e., inappropriate disclosure of data containing personally identifiable information (PII). One dimension of measurement of privacy harm is to assess the degree of clarity of different notices about the terms of data disclosure and use. Misuse and abuse of data are likely covert activities, discovered by investigators and journalists. Effectively tracking of data breaches was not possible until law(s) required that holders of data disclose breaches.

[10] `https://cabforum.org/`

2.4 Harms to innovation, competition, market power, and economic growth

- Innovation. Measurement of innovation uses economic proxies such as rate of sector-specific IPOs. The deeper question is the how different innovations may lead to benefit or harm, including in terms of economic growth. In our harms classification, we observed that not all innovation is pro-consumer and a driver of economic growth. Some innovation may further entrench a dominant incumbent. Characterizing different sorts of innovation is beyond the scope of this paper; we only observe that measurements that attempt to gauge market power and correlate economic growth to micro-economic actions are both difficult and fraught with uncertainty. Similarly, attempting to measure the potential harm that arises on a social media platform when it evolves to make the user experience more "sticky" (in order to keep the user on the platform, e.g., to show them more ads) is speculative and weakly structured.

- Market power. Measurement of market power will require new approaches to accommodate the dynamics of information capitalism, and multi-sided platforms with different degrees of market share on the different sides. Feld's 2019 book making the case for regulation of digital platforms[11] proposed a new metric of market power, called *cost of exclusion*, which maps directly to the potential of harm from market power of digital platforms. The CoE estimates the cost to a third party firm or user due to being excluded from a platform. This simple framing avoids the need to directly assess market power, or whether a platform is multi-sided. However, like innovation-related harms, estimation of CoE is neither straightforward nor free of contention about assumptions.

2.5 Harms to journalism, the marketplace of ideas, and the political processes

Our analysis of harms[12] acknowledged the existence of high-level societal harms, which arise not from the specifics of the Internet architecture, but from more general and fundamental characteristics of the digital ecosystem—such network effects that drive toward consolidation. Feld[13] identified remedies to mitigate some of harms that result from these tendencies. Assessing such harms will likely be more subjective here then in the lower layers of the system, and will require interdisciplinary expertise to specify, scope, and interpret measurements, including measures of human behavior.

3 Barriers to data collection by independent researchers

Independent, third-party researchers can only work with the data that is available to them. They have no ability to compel the release of data. Traditionally, the network measurement community has used two methods to gather data: active probing of networks, and passive observation of traffic, which requires deploying monitoring instrumentation at a point in a network where it can observe traffic. Both methods have significant limitations. Active probing can only infer a few aspects of the network being probed. Passive monitoring raises serious concerns about privacy, both of individuals communicating across the network, as well as the operator of the network. There is no incentive for a commercial network operator to let any unaffiliated party gather data from its network. Sometimes it is illegal to do so, but even if legal barriers are overcome, there is always a risk that data related to a provider's service offering can shed light on aspects of that service that the provider wished to keep secret, or might reflect poorly on that operator in some respects. Further, the increasing bandwidth of today's network links has rendered it technically challenging and prohibitively expensive to collect and store network traffic traces for scientific research.

The daunting barriers to collecting data needed for scientific research on Internet infrastructure motivated the National Science Foundation to sponsor two Workshops on Overcoming Measurement Barriers to Internet Research (WOMBIR) in 2021.[14] This section summarizes the barriers discussed at that workshop, which included evolution of industry structure, lack of capital and incentive for (especially longitudinal) data collection, privacy implications, and limitations of current ethical review mechanisms. The final report has more details on these and other topics discussed at the workshop.[15]

[11] Harold Feld. *The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms*. Roosevelt Institute and Public Knowledge, May 2019.

[12] Clark and claffy, "Toward a Theory of Harms in the Internet Ecosystem".

[13] Feld, *The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms*.

[14] *NSF-sponsored Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021) – Parts I and II*. `https://www.caida.org/workshops/wombir/2101/`. 2021.

[15] "NSF-sponsored Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021) – Final Report".

3.1 Evolution of industry shifts traffic off public Internet

Researchers can perform active probing from the edge of the network, especially a public, consumer-facing network, without specific permission of the network operator. However, the set of networks that are accessible in this way is a small fraction of the networks in the Internet, and not where most traffic originates. Less and less of total traffic transits the global public Internet—the part of the Internet ecosystem most amenable to independent measurement. Modern applications typically depend on platform elements in addition to the public Internet, such as cloud facilities and content distribution networks. CDNs operate rich networks of servers, often using anycast or DNS-based network traffic redirection. Such servers are often hosted in third-party networks, partially masking the CDN's presence from observation. It is difficult and sometimes impossible to attach probing points onto those elements. Probing may be prohibited by terms of service, cost, or technical factors. The result is a growing scope of assets on which application designers (and users of those apps) depend that are beyond the scope of independent measurement.

The changing character of industry structure, with increasing concentration, may raise concerns about market power. The centralization of services will have an uncertain impact on network resilience (Section 4.2). In particular, providers of large-scale services with many customers may be able to invest more in mechanisms that improve resilience, but a failure of such a service can disrupt many dependent services.

3.2 Lack of capital and incentive for longitudinal data collection and sharing

Organized collection and curation of data is expensive. Researchers may collect data in support of a specific undertaking, such as a PhD thesis, but sustaining such an effort is typically cost-prohibitive after the student graduates or moves on to another publishable topic. Sustainability costs arise from the size of the Internet and the resulting data sets, and the effort to make the data usefully accessible.

Incentives for publication, funding, and graduation/promotion also favor one-off snapshots that may become stale. Program committees favor novelty over repeated analysis of previous results. Publishing replication studies can be quite challenging, especially if those results have not changed. Evaluation of scientific promotion does not always value artifacts such as data sets or infrastructure.

Structural limitations of funding agencies reinforces these practices. Most funding sources fund short (three-year) research projects, with no mechanism for extending the budget by a small amount to enable sustained measurement. Moreover, funding agencies do not yet have a way to evaluate longitudinal Internet measurement research, nor an explicit program to review and renew longitudinal activities.

Existing attempts to encourage public data and revisiting of results have included community awards, reproducibility badges, and reproducibility tracks at conferences. These have had only partial success due to their low professional impact, relative to promotion, publications, and degrees. What the community can do on its own does not provide sufficient incentive for meaningful change.

3.3 Privacy implications of infrastructure measurements

Measurement data spans a spectrum of identifiability, from personally identifiable information (PII) that includes, e.g., an email address, to information aggregated such that it cannot be related to an identifiable person. Most Internet measurement data lies between these extremes. Common examples are data sets that include source and destination IP addresses, location, and/or portions of packet payloads. The growing importance of safeguarding privacy in the personal-data-driven ecosystem has triggered three trends that increase the complexity of measurement efforts.

1. **Application traffic is increasingly encrypted.** Thus, even when passive collection of application-layer information is feasible, its utility is limited. Observers must infer what they can from data flows without being able to see content.

2. **Evolution of complex privacy law.** Many researchers do not understand the implications of privacy laws and regulations to academic research. In the case of health care in the U.S., HIPAA (the Health Insurance Portability and Accountability Act) has specific provisions that govern research practices that use medical data. Internet infrastructure researchers need similar procedures to protect misuse of PII, and governments must affirm that those procedures are consistent with their laws and regulations.

In: *Computer Communiations Review* (2021). https://www.caida.org/catalog/papers/2021_wombir2021_report/wombir2021_report.pdf.

The most pertinent regulations are the European General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). Although companies may be subject to the GDPR and/or the CCPA, it is unclear that university researchers are.[16] Both the GDPR and the CCPA encourage forms of data minimization such as pseudonymization and de-identification; these developments are increasing interest in disclosure control technologies that can perform such data minimization.

3. **Technological frameworks to support work with PII have emerged,[17] but their utility is not yet clear.** For example, with differential privacy, a researcher does not obtain direct access to a data set but may submit queries; the amount of distortion in the result of the query is calibrated to ensure that a metric of privacy leakage remains below a specified threshold. A critical gap remains: *identifying how to apply these privacy preserving technologies to networking problems,* and where networking questions do not fit – for example, when a few queries would consume the entire privacy budget. Measurement researchers face a steep learning curve in order to leverage these advanced privacy-preserving frameworks. It is generally easier just to collaborate with someone else who has measurement data, even if that person is at a company and the data must remain secret and any results must have approval from company lawyers before publishing.

3.4 Limitations of ethical review institutions

Another barrier to undertaking and sustaining measurement activities is the lack of familiarity (and consistent treatment) of privacy concerns by Institutional Review Boards (IRBs). Institutional Review Boards are tasked with ethical and regulatory oversight of measurement research that involves the collection, use, or sharing of personally identifiable information, consistent with ethical principles, and more recently with privacy laws and regulations.[18] In this rapidly evolving research ecosystem, IRB decisions are surprisingly variable across institutions and the community would benefit from more uniformity.

The IRB model has another limitation for today's Internet research environment. "Curiosity-driven" research is a valuable component of evolving our understanding of the Internet ecosystem. There is an old saying in Computer Science: "Take a measurement, find a bug". But the idea of exploratory, curiosity-driven research is at odds with the constraints of an IRB. IRBs require a well-defined question and experimental protocol before approving a project. Exploratory research does not always start with a well-formed research question. Important insights have resulted from using data to explore questions that were not contemplated when the data was collected. Similarly, regulations such as the GDPR require that the researcher identify and state the purpose for they are collecting data containing PII, before the collection begins, and must commit that they will use the data only for that purpose. These are important and necessary safeguards against the misuse of data, but they inhibit exploratory research with sensitive data sets.

4 Public policy problems that create pressure for change

The scientific research community cannot by itself solve the problems we describe. If they are worth solving, it will require higher-level attention, and it is not clear who has that responsibility. In this section we revisit two public policy issues that evidence suggests are rising to the level where the public interest calls for

[16] The GDPR applies to entities in the European Union, to data processing related to the offering of goods or services to European subjects, and to the monitoring of the behavior of European subjects; see GPDR Recitals 22-24. The CCPA applies to for-profit businesses; see CCPA Section 1798.140(d).

[17] Cynthia Dwork et al. "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography*. Ed. by Shai Halevi and Tal Rabin. 2006. ISBN: 978-3-540-32732-5; CACM. "Differential Privacy: The Pursuit of Protections by Default". In: *Commun. ACM* 64.2 (Jan. 2021). ISSN: 0001-0782. DOI: 10.1145/3434228. URL: https://doi.org/10.1145/3434228; David Evans, Vladimir Kolesnikov, and Mike Rosulek. "A Pragmatic Introduction to Secure Multi-Party Computation". In: *Foundations and Trends® in Privacy and Security* 2 (2018). free version at http://securecomputation.org/. URL: http://dx.doi.org/10.1561/3300000019; Yehuda Lindell. "Secure Multiparty Computation". In: *Comm. ACM* (Dec. 2020). DOI: 10.1145/3387108. URL: https://doi.org/10.1145/3387108; Paul Francis, Sebastian Eide, and Reinhard Munz. "Diffix: High-Utility Database Anonymization". In: *Annual Privacy Forum*. June 2017, pp. 141–158. ISBN: 978-3-319-67279-3. DOI: 10.1007/978-3-319-67280-9_8; Henry Corrigan-Gibbs and Dan Boneh. "Prio: Private, Robust, and Scalable Computation of Aggregate Statistics". In: *USENIX Conference on Networked Systems Design and Implementation*. NSDI'17. Boston, MA, USA, 2017.

[18] Kenneally, Erin and Dittrich, David. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. http://ssrn.com/abstract=2445102. 2012; Dittrich, David and Kenneally, Erin and Bailey, Michael. "Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report". In: http://ssrn.com/abstract=2342036. 2013.

orchestrated attention to the problem and its mitigation: Internet access, and security and resilience of the infrastructure. Both issues are receiving increasing attention from governments, suggesting an impending point of transition. We suggest the policy and scientific research community should recognize this inflection point and help shape it in constructive ways.

4.1 Understanding deployment and uptake of Internet access

Understanding the state of Internet access is a grand challenge because it inherently comes with grand scale, and deep societal importance. Technologies such as 5G and low-earth-orbit satellite expand the range of options for access, as well as the range of performance and affordability of these options. Access properties of interest include deployment coverage, availability, adoption, throughput, latency, reliability, and usage. Some of this data is notoriously hard to acquire, and public debate on the accuracy of these data sets for quantifying differences in these properties across the country (the digital divide) has continued for decades.

In part due to the FCC's role in promoting broadband deployment, there are many existing data sets related to broadband. (Section 1.1). The federal government's Data Catalog indexes over 600 broadband data sets,[19] including measurements since 2011 from the FCC's Measuring Broadband America (MBA) program, including sample-based views of throughput, latency, jitter, DNS performance of U.S fixed broadband access services. This longitudinal data set has seen substantive but limited use by researchers in their study of U.S. broadband.[20] Understanding longitudinal trends in access properties require creative and technically sound methods to use all forms of data collection, even those that contain inaccuracies.

As the need to ensure access to a specified level of broadband service increases, maximizing utility of existing data sets will require federation of data collection, standardization of reporting, methods to overcome measurement bias (e.g., from crowdsourced measurements), multi-level spatial analysis and representation, and support for local contributions to national data sets that preserve privacy. Moreover, effectively mapping broadband access over time requires measurements that go beyond basic access, to quality of service, reliability, and affordability.

4.2 Understanding Internet security and resilience

The biggest driver today for increased collection of Internet data is security, but given the Internet's status as critical infrastructure underlying (and in many cases controlling) other critical infrastructure, the more general goal of *resilience* is a growing societal concern. Abstractly, the resilience of a system is the ability to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation, whether due to malice, operator error, hardware or software faults, or any other reason. But this definition says nothing about how to achieve resilience. The real measure of resilience is whether the user experience is disrupted, but user-level impairment is sufficiently hard to measure that researchers use proxy measures, such as "bit-risk miles", a technical measure of how much capacity and connectivity is lost with a given failure.[21] The relationship between these proxy measures and impairment is unclear.

We suggest another framing, rooted in analysis of the three key systems that must function for the Internet to provide services: the routing system, the naming system, and the Certificate Authority system. Each system has its own approach to providing resilience, and these systems have interdependencies. Thus, the overall resilience of the Internet is a complex, multi-dimensional space that is difficult to assess.

[19] U.S. General Services Administration (GSA). *Broadband Data Resources in data.gov*. `https://catalog.data.gov/dataset?q=broadband`. 2021.

[20] Zachary S. Bischof, Fabián E. Bustamante, and Rade Stanojevic. "Need, Want, Can Afford – Broadband Markets and the Behavior of Users". In: *Proceedings of the ACM Internet Measurement Conference*. `https://doi.org/10.1145/2663716.2663753`. 2014; John P. Rula, Zachary S. Bischof, and Fabian E. Bustamante. "Second Chance: Understanding Diversity in Broadband Access Network Performance". In: *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*. C2B(1)D '15. London, United Kingdom: Association for Computing Machinery, 2015, pp. 9–14. ISBN: 9781450335393. DOI: `10.1145/2787394.2787400`. URL: `https://doi.org/10.1145/2787394.2787400`; Zachary S. Bischof, Fabian E Bustamante, and Rade Stanojevic. "The utility argument – making a case for broadband SLAs". English (US). in: *Passive and Active Measurement - 18th International Conference, PAM 2017, Proceedings*. Ed. by Steve Uhlig, Johanna Amann, and Mohamed Ali Kaafar. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2017, pp. 156–169. ISBN: 9783319543277. DOI: `10.1007/978-3-319-54328-4_12`.

[21] Brian Eriksson, Ramakrishnan Durairajan, and Paul Barford. "Riskroute: A framework for mitigating network outage threats". In: *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. 2013, pp. 405–416.

Routing System. The routing system must be working in order for the routers to forward packets. Both the routing protocols used internally to ISPs and the global routing protocol (Border Gateway Protocol or BGP) compute routes dynamically, and will fall back to alternative routes if they exist when one route fails. Researchers have run simulations to predict the loss of connectivity from a given failure. However, such research can only explore the first tier of resilience: the alternative routes that are announced via BGP. In response to failures, network operators can change their routing policies and enable new paths, and they can add new physical connectivity to change the network topology. Network operators can rapidly perform such physical changes that are local to a data center. This agility makes is difficult to measure or characterize the resilience of the Internet to failures of links and routers.

Domain Name System. The Domain Name System must be working so that users can translate domain names into Internet addresses. Again, there are tiers of mechanism that add resilience to the DNS. Local name resolvers cache the results of queries to avoid having to query authoritative parts of the system. Large DNS resolvers that answer queries about the top-level names are highly replicated, sometimes with heterogeneous implementations, and in many cases exploiting *anycast* routing (where one IP address is assigned to multiple distributed nodes), which enables resilience in case of resolver failure. These tiers of mechanisms make it difficult to estimate how a failure of a resolver will affect the user experience.

Certificate Authority system. The Certificate Authority system must be available so that users can verify that they have reached the intended web site (or other service). The CA system is complex, and less mature than the DNS. Many advisories describe how to make a given CA resilient, but we find less discussion about how the overall system can enhance its resilience.

The mechanisms that ensure resilience of these systems are often not active when the Internet is operating normally, which renders elusive the capability to assess resilience of these systems. Operators can do chaos engineering to assess how their systems respond to failures, but third-party researchers cannot take that approach. In this circumstance, a research agenda must be opportunistic – leveraging sources of data that that shed light on aspects of resilience. Case studies of outages (which can represent failures of resilience) illustrate what did *not* work. It is harder to measure the "near misses" as the airlines do – when did mechanisms kick in to preserve the quality of the user experience? A first step must be to review existing literature of attempts to measure various aspects of resilience of Internet infrastructure – what aspect of resilience they studied, using what data, and what methods. The goal would be to generalize from existing studies and develop a conceptual overview of aspects of resilience.

5 Concrete recommendations and next steps

In this section we consider recently proposed concrete steps to make progress on the gap in empirical data to inform policymaking, and the role of academics in doing so.

5.1 Systematically analyze risks of sharing proprietary data

Sharing of data by the private sector carries risk beyond concerns about inappropriate release of PII. Release of certain data could lead to adverse commentary on some stakeholder, or policies adverse to the stakeholder's interests, and these concerns triggers understandable hesitation. If we can understand more about the structure of proprietary data, we may be able to improve the options for controlled access for research purposes. Even if sharing such data requires regulatory support, understanding these risks is prerequisite to developing reasonable regulations. We identify three sorts of proprietary data, with different barriers to sharing.

- **Functional data.** Functional data arises from the business practices of the enterprise. In general, functional data is unique to a given firm. Examples include the network of friend relationships in Facebook or the retweet structure of Twitter. Such data is valuable to researchers attempting to understand the propagation of disinformation. Other examples include a DNS registrar's database of metadata regarding ownership of domain names, business and technical aspects of interconnection agreements among ISPs (which relate to how traffic is routed and measures of resilience), and packet flow data.

There are obvious counter-incentives to sharing functional data, including the operational complexity of making it externally available, and the risk that it reveals proprietary information, Misuse of the data may violate the terms under which the firm acquired it. This latter applies particularly to Personally Identifiable Information (PII). Since release of such data represents a risk to the firm, the firm will need to control data disclosure.

- **Event data.** Event data relates to things that happen to firms, including penetrations, exfiltrations, financial losses, etc. Event data is in general not unique to an enterprise. All enterprises are attacked, or may suffer a breach or a loss. Data of this sort can inform a range of research, including evolving patterns of attack or losses. Firms may be reluctant to reveal firm-specific event data, but may benefit from industry-wide aggregation of such data. For this class of data, it may be possible to define general practices that are sufficient to reduce risk (Section 3.3). Geoff Huston recently praised Akamai[22] for their unusually full and careful reporting of a recent outage they suffered, an outage visible enough to get press coverage.

- **Observations.** Observations are data explicitly gathered to inform properties of security or performance, e.g., blocklists, collections of malware, speed tests. When proprietary, the owner will likely sell such data to firms that want to protect themselves or their customers. Researchers can often use historic forms of such data with little commercial value (customers want real-time threat data to react to it), but high research value. An untapped opportunity is to archive historical data points of certain commercial data sets, in cases where archives can support non-commercial longitudinal research on security and resilience, without interfering with the revenue model for the data.

These categories are not rigid or exclusive. Some firms may translate some of their functional data into observations they sell, which may further inhibit their interest in sharing data with researchers. With increasing interest in the security and resilience of the Internet, we can expect increasing calls for disclosure of event data, along the lines of the data breach laws. Firms will be more willing to accept a requirement for disclosure if they understand the terms under which the disclosed data will be used, and the steps taken to reduce risk. Observations today are sometimes released to researchers on a no-cost basis, under the assumption that the results may benefit the firm as well as society, and the commercial risk is low. But these arrangements are typically one-time events, which makes replication of the research difficult. The government could purchase the data with the understanding that only vetted researchers will use it in approved ways, and then make it available to those vetted researchers.

It is important to learn how other fields of science and critical infrastructure research have addressed their data challenges.

5.2 Potential roles for academics in proposed approaches

Other parts of the globe are moving to regularize cybersecurity data, and they have explicitly recognized the importance of engaging and sustaining the academic research establishment in this effort. A recent announcement from the European Union illustrates that this area is receiving substantial attention. In their proposed regulation for Digital Services,[23] they discuss the importance of ensuring access to proprietary data by the academic research community. The report states:

> *"Investigations by researchers on the evolution and severity of online systemic risks are particularly important for bridging information asymmetries and establishing a resilient system of risk mitigation, informing online platforms, Digital Services Coordinators, other competent authorities, the Commission and the public. This Regulation therefore provides a framework for compelling access to data from very large online platforms to vetted researchers.*
>
> *... In order to be vetted, researchers shall be affiliated with academic institutions, be independent from commercial interests, have proven records of expertise in the fields related to the risks investigated or related research methodologies, and shall commit and be in a capacity to preserve the specific data security and confidentiality requirements corresponding to each request."*

[22] Geoff Huston. *Opinion: Why is this unusual?* `https://blog.apnic.net/2021/07/27/opinion-why-is-this-unusual/`. July 2021.

[23] European Commission. *Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services*. `https://eur-lex.europa.eu/legal-content/`. 2020.

This regulation emphasizes a structure that enables and encourages the academic community to work with proprietary data, sending an important signal that they intend to make their academic research establishment a recognized part of shaping the future of the Internet in the European Union.

The U.S. has not yet taken such a pro-active stance. The best evidence of U.S. consideration of this issue is in the U.S. Cyber Solarium Commission report from 2019. That report set out a strategic plan to improve the security of cyberspace.[24] Among its many recommendations is that the government establish a U.S. Federal Bureau of Cyber Statistics (BCS), to provide the government with the information that it needs for informed planning and action. A recent report from the Aspen Institute echoed this call.[25] Legal academics and lobbyists have already started to consider its structure.[26] There have thus far been no technical, scientific, or engineering voices in this conversation. The Solarium report provides an opportunity to consider the relationship academics could or should have with such a government function.[27]

In this context, the Solarium report makes two concerning observations. First, the report charges the Bureau with the task of "purchasing private or proprietary data repositories," implying that much of the data will not be public. Second, the report proposes that the Bureau "host academics as well as private-sector and independent security researchers as a part of extended exchanges". There is no consideration of engagement of on-campus research groups including graduate students, or release of datasets that can stimulate independent development of advanced analytical techniques such as machine learning. Creation of a Bureau of Cyber Statistics could have the unintended consequence of sidelining the academic research community that has been measuring and analyzing the Internet, and which can provide unbiased advice to the nation about the societal impacts of the Internet, as well as its technical characteristics.

5.3 Potential roles for research funding agencies in proposed approaches

In addition to the European Union's above governments could undertake other steps that would narrow the empirical gap that Internet policymakers face.

1. The government could help **navigate misalignment of incentives that impede data sharing to support scientific research**. This includes shepherding data use agreements with providers to facilitate industry contribution of large, shareable data sets for research and STEM work force training. As an example, for over 10 years, DHS supported the IMPACT project to promote sharing of data for cybersecurity research years,[28] and published the legal agreements governing data sharing so others could benefit.[29] Another output of these programs was the Menlo Report,[30] which proposed concrete risk assessment methods and tools for sharing information security incident and threat data. Inspired by the 1979 Belmont Report[31] supporting ethical research in medical and behavioral sciences, the Menlo Report emphasized the following principles: identification of stakeholders and informed consent; balancing risks and benefits; fairness and equity; and compliance, transparency and accountability,

2. To navigate the entire data management lifecycle, funding agencies could also **facilitate standardization of data practices** by promoting (funding) the creation of working groups to standardize rules of data set generation and sharing of data artifacts, and to create common application platforms and tooling for maintaining and sharing best-practice pipelines for issuing, processing, and publishing measurements.

3. The government could fund demonstrations of compensating benefits to the private sector in a program of data sharing. Each actor in the Internet ecosystem may have an accurate view of their part of the

[24] *Cyberspace Solarium Commission report*. https://www.solarium.gov/report. 2020.

[25] The Aspen Institute. "A National Cybersecurity Agenda for Resilient Digital Infrastructure". In: (2020). https://www.aspeninstitute.org/longform/a-national-cybersecurity-agenda-for-resilient-digital-infrastructure/.

[26] Chas Kissick and Paul Rosenzweig. "Considerations for the Structure of the Bureau of Cyber Statistics". In: *Lawfare* (Oct. 2020). https://www.lawfareblog.com/considerations-structure-bureau-cyber-statistics.

[27] National Academy of Sciences. *Principles and Practices for a Federal Statistical Agency, Edition 7*. https://www.nap.edu/resource/25885/P\&P\%207\%20Higlights.pdf.

[28] Department of Homeland Security. *IMPACT Cybertrust Data Sharing Project*. https://www.impactcybertrust.org/.

[29] Department of Homeland Security. *IMPACT Cybertrust Legal Tools*. https://www.impactcybertrust.org/tools_legal.

[30] Kenneally, Erin and Dittrich, David, *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*; Dittrich, David and Kenneally, Erin and Bailey, Michael, "Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report".

[31] Office of the Secretary, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. "Ethical Principles and Guidelines for the Protection of Human Subjects of Research". In: (Apr. 1979). https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

- Data is made available in curated repositories, or otherwise provided in ways that allows adequate access for legitimate scientific research
- Access requires registration with data source and legitimate research need
- Standard anonymization methods are used where needed
- Recipients agree to not repost corpus
- Recipients agree that they will not deanonymize data
- Recipients can publish analysis and data examples necessary to review research
- Recipients agree to use accepted protocols when revealing sensitive data, such as security vulnerabilities or data on human subjects
- Recipients agree to cite the repository and provide publications back to repository
- Repository can curate enriched products developed by researchers
- Governments should affirm that sharing data under such standard usage agreements is encouraged and is not a legal liability.
- Funding agencies, journals, conferences and professional societies should encourage research conducted under these conditions.

Table 1: *Such codes of conduct enable responsible sharing of data in ways that protect stakeholders while allowing research. To keep science and engineering communities competitive, governments will need to encourage and incentivize sharing data under such standard usage agreements. Funding agencies, journals, conferences and professional societies should incentivize research conducted under these conditions.*

system, but not about the state of their competitors, or the larger ecosystem. Allowing neutral third parties to obtain data from multiple actors can give the private sector, as well as governments and society, a global view of the state of the Internet. But the government will have to find ways to limit liability as a result of responsible sharing of data for documented scientific research.

4. The government could also encourage/promote specific sharing of data to support academic training of STEM professionals to work with large realistic Internet data sets. While synthetic network data can be used for classroom exercises, serious research of the sort that leads to professional development requires real data, with the genuine potential for new discovery.

5. The U.S. government could send a strong signal to the private sector that builds and operates the Internet: data sharing is a necessary aspect of sustaining critical infrastructure, the Internet has now reached this level of maturation, and (as is true in other aspects of society) responsible data sharing needs to be part of normal practice. Developing this model now is a worthwhile activity before some future Internet catastrophe forces an ad-hoc approach to Internet data sharing that would be less beneficial to operators, policymakers, and citizens.

6. The federal government can contribute to advancing the application of privacy-preserving techniques to Internet infrastructure data, by promoting cross-fertilization among the fields of Internet measurement, privacy-preserving algorithms, and privacy laws and regulations.

7. Because technical privacy-preserving tools will not resolve all concerns about the sharing of sensitive data, the government could also promote well-understood practices, used in this and other sectors, to responsibly share data with qualified independent scholars (Table 1) to allow replication or extension of previous work.

8. Governments could promote the creation and operation of an **oversight committee** (a kind of meta-IRB) to oversee community measurement platforms, develop **best practices** around data anonymization, and **support matchmaking** between researchers and data providers. The National Science

Foundation may be best positioned to undertake such an effort. In the U.S., university IRBs exist as part of a set of requirements to receive federal funding, so the government could use this rubric to enrich IRBs understanding of best practices associated with Internet research, of privacy preserving techniques, and of approaches articulated in privacy laws and regulations sufficiently to evaluate privacy risks, even if university research is not subject to those regulations.

6 Final Thoughts

There are limits to what any given community or even set of stakeholders can do to overcome barriers to Internet measurement in the public interest. There are many actors in the ecosystem–researchers and the academic context within which they sit (with its priorities for publication, funding, advancement and tenure), service providers, governments, advocates for various objectives ranging from privacy to improved access, and funding agencies. Changing the landscape of network measurement would require adjustment in many parts of this ecosystem. We recognize the argument that the benefit does not justify the cost. But we also recognize that the Internet is the only critical infrastructure without dedicated government oversight, including a data collection function. It would surprise us if this lack of oversight persisted for another decade. In our view, the question is not whether there needs to be measurement of the Internet in the public interest. The question is how to sustainably achieve it.

References

Anderson, Ross and Barton, Chris, and Bohme, Rainer, and Clayton, Richard, and van Eeten, Michel and Levi, Michael, and Moore, Tyler and Savage, Stefan. "Measuring the Cost of Cybercrime". In: *Workshop on Economics and Information Science*. https://www.econinfosec.org/archive/weis2012/papers/Anderson_WEIS2012.pdf. 2012.

Bischof, Zachary S., Fabián E. Bustamante, and Rade Stanojevic. "Need, Want, Can Afford – Broadband Markets and the Behavior of Users". In: *Proceedings of the ACM Internet Measurement Conference*. https://doi.org/10.1145/2663716.2663753. 2014.

Bischof, Zachary S., Fabian E Bustamante, and Rade Stanojevic. "The utility argument – making a case for broadband SLAs". English (US). In: *Passive and Active Measurement - 18th International Conference, PAM 2017, Proceedings*. Ed. by Steve Uhlig, Johanna Amann, and Mohamed Ali Kaafar. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2017, pp. 156–169. ISBN: 9783319543277. DOI: 10.1007/978-3-319-54328-4_12.

Brian Cute. *A conversation about evolving the effectiveness of our multistakeholdermodel*. https://www.icann.org/en/system/files/files/draft-evolving-multistakeholder-model-issues-list-25apr19-en.pdf. Mar. 2019.

CACM. "Differential Privacy: The Pursuit of Protections by Default". In: *Commun. ACM* 64.2 (Jan. 2021). ISSN: 0001-0782. DOI: 10.1145/3434228. URL: https://doi.org/10.1145/3434228.

Chas Kissick and Paul Rosenzweig. "Considerations for the Structure of the Bureau of Cyber Statistics". In: *Lawfare* (Oct. 2020). https://www.lawfareblog.com/considerations-structure-bureau-cyber-statistics.

claffy, k., G. Polyzos, and H. Braun. "Application of sampling methodologies to wide-area network traffic characterization". In: *ACM SIGCOMM*. May 1993.

Clark, D. and k. claffy. "Toward a Theory of Harms in the Internet Ecosystem". In: *Telecommunications Policy Research Conference (TPRC)*. Sept. 2019.

Corrigan-Gibbs, Henry and Dan Boneh. "Prio: Private, Robust, and Scalable Computation of Aggregate Statistics". In: *USENIX Conference on Networked Systems Design and Implementation*. NSDI'17. Boston, MA, USA, 2017.

Cyberspace Solarium Commission report. https://www.solarium.gov/report. 2020.

Department of Homeland Security. *IMPACT Cybertrust Data Sharing Project*. https://www.impactcybertrust.org/.

— *IMPACT Cybertrust Legal Tools*. https://www.impactcybertrust.org/tools_legal.

Dittrich, David and Kenneally, Erin and Bailey, Michael. "Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report". In: http://ssrn.com/abstract=2342036. 2013.

Dwork, Cynthia et al. "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography*. Ed. by Shai Halevi and Tal Rabin. 2006. ISBN: 978-3-540-32732-5.

Eriksson, Brian, Ramakrishnan Durairajan, and Paul Barford. "Riskroute: A framework for mitigating network outage threats". In: *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*. 2013, pp. 405–416.

European Commission. *Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services*. https://eur-lex.europa.eu/legal-content/. 2020.

Evans, David, Vladimir Kolesnikov, and Mike Rosulek. "A Pragmatic Introduction to Secure Multi-Party Computation". In: *Foundations and Trends® in Privacy and Security* 2 (2018). free version at http://securecomputation.org/. URL: http://dx.doi.org/10.1561/3300000019.

Federal Communications Commission. *The National Broadband Plan: Connecting America*. http://download.broadband.gov/pl broadband-plan.pdf. 2010.

Feld, Harold. *The Case for the Digital Platform Act: Market Structure and Regulation of Digital Platforms*. Roosevelt Institute and Public Knowledge, May 2019.

Francis, Paul, Sebastian Eide, and Reinhard Munz. "Diffix: High-Utility Database Anonymization". In: *Annual Privacy Forum*. June 2017, pp. 141–158. ISBN: 978-3-319-67279-3. DOI: 10.1007/978-3-319-67280-9_8.

Huston, Geoff. *Opinion: Why is this unusual?* https://blog.apnic.net/2021/07/27/opinion-why-is-this-unusual/. July 2021.

Internet Corporation for Assigned Names and Numbers. *Board Action on Competition, Consumer Trust, and Consumer Choice Review*. https://www.icann.org/news/blog/board-action-on-competition-consumer-trust-and-consumer-choice- review. Mar. 2019.

— *Board Action on Security, Stability, and Resilience 2 Review*. https://www.icann.org/en/system/files/bm/rationale-ssr2-22jul21-en.pdf. July 2021.

— *ICANN Articles of Incorporation*. https://www.icann.org/resources/pages/governance/articles-en.

— *ICANN Bylaws*. https://www.icann.org/resources/pages/governance/bylaws-en/.

Kenneally, Erin and Dittrich, David. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. http://ssrn.com/abstract=2445102. 2012.

Kleinig. "Crime and the Concept of Harm". In: *American Philosophical Quarterly* 15.1 (1978).

Lindell, Yehuda. "Secure Multiparty Computation". In: *Comm. ACM* (Dec. 2020). DOI: 10.1145/3387108. URL: https://doi.org/10.1145/3387108.

National Academy of Sciences. *Principles and Practices for a Federal Statistical Agency, Edition 7*. https://www.nap.edu/resource/25885/P\&P\%207\%20Higlights.pdf.

"NSF-sponsored Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021) – Final Report". In: *Computer Communiations Review* (2021). https://www.caida.org/catalog/papers/2021_wombir2021_report/wombir2021_report.pdf.

NSF-sponsored Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021) – Parts I and II. https://www.caida.org/workshops/wombir/2101/. 2021.

Office of the Secretary, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. "Ethical Principles and Guidelines for the Protection of Human Subjects of Research". In: (Apr. 1979). https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

Rula, John P., Zachary S. Bischof, and Fabian E. Bustamante. "Second Chance: Understanding Diversity in Broadband Access Network Performance". In: *Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data*. C2B(1)D '15. London, United Kingdom: Association for Computing Machinery, 2015, pp. 9–14. ISBN: 9781450335393. DOI: 10.1145/2787394.2787400. URL: https://doi.org/10.1145/2787394.2787400.

The Aspen Institute. "A National Cybersecurity Agenda for Resilient Digital Infrastructure". In: (2020). https://www.aspeninstitute.org/longform/a-national-cybersecurity-agenda-for-resilient-digital-infrastructure/.

U.S. Congress. *Telecommunications Act of 1996, Pub. L. No. 104-104, 706(a), 110 Stat. 56, 153 (codified as 47 U.S.C. 1302)*. 1996.

U.S. General Services Administration (GSA). *Broadband Data Resources in data.gov*. https://catalog.data.gov/dataset?q=broadband. 2021.