



# Digital Equity 2.0: How to Close the Data Divide

---

By Gillian Diebold and Daniel Castro | May 8, 2023

The United States has strived to address digital inequities in the Internet economy through programs that combat the “digital divide.” But in the data economy, a number of social and economic inequities arise from a lack of data collection or use of data. These inequities—the “data divide”—require new policy solutions to ensure that all Americans are represented in data and can put it to use. To tackle this new challenge, policymakers face two starkly different options: Option A) hold back data collection and data-driven technologies until they are equitable for everyone; or option B) allow the data-driven technologies to prosper while working to increase access for everyone. To close the digital divide, U.S. policymakers have chosen option B. But for the data divide, many are flirting with option A, when they should be choosing option B.

## INTRODUCTION

For the last decade, closing the digital divide, or the gap between those subscribing to broadband and those not subscribing, has been a top priority for policymakers. But high-speed Internet and computing device access are no longer the only barriers to fully participating and benefiting from the digital economy. Data is also increasingly essential, including in health care, financial services, and education. Like the digital divide, a gap has emerged between the data haves and the data have-nots, and this gap has introduced a new set of inequities: the data divide.

Policymakers have put a great deal of effort into closing the digital divide, and there is now near-universal acceptance of the notion that obtaining

---

widespread Internet access generates social and economic benefits. But closing the data divide has received little attention. Moreover, efforts to improve data collection are typically overshadowed by privacy advocates' warnings against collecting any data. In fact, unlike the digital divide, many ignore the data divide or argue that the way to close it is to collect vastly less data.<sup>1</sup> But without substantial efforts to increase data representation and access, certain individuals and communities will be left behind in an increasingly data-driven world.

This report describes the multipronged efforts needed to address digital inequity. For the digital divide, policymakers have expanded digital connectivity, increased digital literacy, and improved access to digital devices. For the data divide, policymakers should similarly take a holistic approach, including by balancing privacy and data innovation, increasing data collection efforts across a wide array of fronts, enhancing access to data, improving data quality, and improving data analytics efforts. Applying lessons from the digital divide to this new challenge will help policymakers design effective and efficient policy and create a more equitable and effective data economy for all Americans.

### The Need for a Data-Rich Society

Data leads to better understanding and decision-making among individuals, businesses, and government. Individuals use data to make better decisions about everything from what they buy to how they plan for the future. Businesses use data to find new customers, automate processes, develop and improve products and services, and inform business decisions. Government agencies use data to cut costs, improve social services, and keep citizens safe. A data-rich society brings benefits in a broad range of areas, as shown in Table 1. To ensure all Americans receive these benefits, policymakers should commit to closing the data divide.

**Table 1: Ten examples of the benefits of data-driven innovation**

| Domain                | Use Case  |
|-----------------------|---|
| <b>Public Health</b>  | The U.S. Centers for Disease Control uses social network analysis to better understand and stem the spread of communicable diseases.              |
| <b>Education</b>      | School teachers and administrators provide preemptive interventions for students at risk of falling behind with the help of predictive analytics. |
| <b>Transportation</b> | The city of Austin, Texas, uses real-time traffic data to time light changes, reduce congestion, and quantify emissions.                          |

---

| Domain               | Use Case  |
|----------------------|---|
| <b>Environment</b>   | Scientific researchers use satellite images to monitor forming crevasses in glaciers and predict future rises in sea levels.  |
| <b>Public Safety</b> | New York City's Fire Department prioritizes inspections based on risk assessments derived from building data, which has resulted in the city reducing the number of annual fire deaths to the lowest since recordkeeping began in 1916. |
| <b>Retail</b>        | User reviews on sites such as Amazon or Yelp help consumers discover the products and retailers they like best.   |
| <b>Government</b>    | The U.S. Securities and Exchange Commission analyzes reporting data from publicly traded companies to identify suspicious filings and inform fraud investigations.  |
| <b>Energy</b>        | Wind energy companies use complex weather models to determine the optimal locations for their turbines.   |
| <b>Manufacturing</b> | Amazon uses predictive modeling on inventory data to coordinate manufacturing supply changes.   |

## POLICIES TO CLOSE THE DIGITAL DIVIDE

State and federal agencies have been working for more than a decade to support the deployment of high-speed broadband. These efforts focus on three core areas: digital connectivity, digital skills, and digital devices. Fifteen federal agencies administer 133 programs to combat the digital divide.<sup>2</sup> Some have broadband access as their main purpose, others as one purpose (of many), and others include broadband access as an ancillary purpose.

Since the 1990s, concerns about gaps in access and use of the Internet and Internet-connected technologies has persisted in policy and scholarly discussions.<sup>3</sup> Taking note of the vast benefits of broadband and potentially growing inequities, Congress directed the Federal Communications Commission (FCC) in 2009 to develop a national strategy to ensure all Americans have “access to broadband capability.”<sup>4</sup> In 2010, the FCC released the National Broadband Plan to outline the ways government can influence the broadband ecosystem. This plan includes establishing policies for competition, innovation, and consumer welfare, along with policies concerning the allocation of the various infrastructure needed for broadband access. The National Broadband Plan also looks specifically at

---

the importance of broadband to key sectors, such as education, health care, and government operations.<sup>5</sup>

The telecommunications landscape has evolved significantly in the decade-plus since the plan's release. In 2010, roughly 65 percent of Americans subscribed to broadband.<sup>6</sup> As of 2019, 90 percent of households subscribed.<sup>7</sup> The understanding that enhanced broadband access serves the public interest was critical for this positive evolution. In a more recent effort to promote digital equity, the Lewis Latimer Plan for Digital Equity and Inclusion from the National Urban League, a nonpartisan civil rights organization, discusses the view that “digital exclusion compounds inequities for historically marginalized groups,” a view that the public and private sectors alike have adopted.<sup>8</sup> The Latimer Plan stresses the importance of deploying networks everywhere, getting near-universal connectivity, and effectively utilizing networks to deliver essential services.

The federal government has supported these goals, as evidenced by a major influx in investment for broadband programs. The 2021 Infrastructure Investment and Jobs Act (IIJA), also referred to as the Bipartisan Infrastructure Law, is the largest investment in broadband deployment and adoption in U.S. history.<sup>9</sup> The National Telecommunications and Information Administration (NTIA), the FCC, states, and other relevant agencies will oversee this funding. A 2022 inventory of federal broadband programs by the Government Accountability Office details the supply of funding that exists for such programs—more than \$48.9 billion for programs in which broadband is the main purpose.<sup>10</sup> That number even excludes new endeavors from 2020 and 2021 created by the CARES Act, Consolidated Appropriations Act of 2021, American Rescue Plan Act of 2021, the IIJA, and other relevant new programs.<sup>11</sup>

The programs supported by this funding influx target digital connectivity, digital literacy, and devices. Many of the programs, particularly those created since 2020, include multiple focus areas, and most of them also include “affordability” in their descriptions.<sup>12</sup> Including affordability in a program's purpose means the program has a targeted scope to include low-income individuals and communities. Funding for affordability can come through discounts, reimbursements, or grant programs. (See appendix for more details on these programs.)

## **POLICIES TO CLOSE THE DATA DIVIDE**

Just like the digital divide hinders certain populations from accessing broadband services, the data divide hinders certain populations from benefiting from data-driven services. But while the digital divide has received a large influx of funding and interest in recent years, few policy efforts have addressed the data divide.

The divide between the data haves and data have-nots will continue to grow unless closing the data divide becomes a key policy priority for those

---

promoting digital equity. Because of a continued lack of equitable data collection, data-driven services do not benefit some individuals and communities. While the digital divide has received an abundance of support and resources from policymakers, data is still largely seen as a one-off input, rather than a core element of digital infrastructure.

In 2014, the Center for Data Innovation first highlighted how individuals and communities lacking high-quality data are at risk of falling behind in an increasingly data-driven world.<sup>13</sup> In the years since, the data economy has been developing at a fast pace, transforming classrooms and hospitals, and enabling better public safety and environmental monitoring. But a number of social and economic inequities can result from a lack of data collection or use of data, and these inequities—the data divide—mean that data-driven services don’t work for certain individuals and communities.<sup>14</sup> Despite the introduction of the Federal Data Strategy (FDS) to accelerate the use of data in the public interest, major data gaps remain, and updates on FDS implementation have been limited.<sup>15</sup> Moreover, a 2022 Center for Data Innovation report finds that little progress has been made to close the data divide, with gaps affecting federal statistics and even entire data systems in health care, education, and financial services.<sup>16</sup>

Closing the data divide will require a holistic approach that addresses a new set of challenges associated with the data economy. Policymakers should also investigate ways to bolster acceptance of collecting and sharing data for social good. Digital divide programs have succeeded in no small part because of the universal notion that broadband access is in the public interest. They should apply that same view to the data divide.

Although there is no one-size-fits-all solution to overcoming digital inequities, crafting a policy response that addresses critical gaps in data quantity, data access, and data quality while reframing privacy policy debates is a strong start. First and foremost, policymakers need to rethink their concept of privacy and accept data as a fundamental enabler of social good. The United States needs more data, but that must go hand in hand with national data privacy legislation. Increasing data collection through better device distribution and usage will ensure that more Americans can reap the benefits of data-driven innovation. Eliminating data minimization clauses will also help ensure that no individual or community experiences data poverty. By enhancing access to data through data portability, open data policies, and open APIs, more communities will be able to participate in the data economy. Moreover, policymakers must think about increasing the representation of historically underserved groups.

### **Create Data-Friendly Privacy Regulations**

Privacy activists routinely argue that collecting, sharing, and using personal data without affirmative consent violates an individual’s human rights. These activists ignore both individual and societal benefits from data. For

---

example, individuals can benefit from better health care if their doctors have better information about them and society can benefit from the development of more effective medical treatments when everyone contributes health data for research. But privacy activists argue that these collective benefits are secondary to individual privacy rights. As a result, privacy laws and regulations have an anti-data bias that impedes beneficial uses of data even in cases where there are minimal privacy risks but positive societal benefits. For example, U.S. federal privacy laws prevent government agencies from combining data about student loan recipients to help future borrowers understand their ability to pay back college loans.<sup>17</sup>

Policymakers should reconsider privacy laws and regulations that may perpetuate the data divide.<sup>18</sup> For example, in Illinois, the Biometric Information Privacy Act prohibits the collection of biometric data without first obtaining informed consent.<sup>19</sup> To avoid running afoul of this law, companies have had to restrict consumers in the state from using popular products and services that use consumer data.<sup>20</sup> Or consider how the Federal Privacy Act of 1974 and other regulations restrict the U.S. Department of Housing and Urban Development from reporting on-time rental payments to the credit bureaus without prior consent from the individual, even when reporting of rental history can enhance accurate reporting of credit scores for public housing tenants.<sup>21</sup>

The European Union's General Data Protection Regulation (GDPR) significantly limits the collection and use of personal data.<sup>22</sup> For example, it requires firms to only collect the minimum amount of data necessary for a specific purpose and not use it for any other purposes without prior consent from the user.<sup>23</sup> In effect, the GDPR prohibits firms from doing anything new and innovative with data. The GDPR also requires that firms undertake a privacy impact assessment for any "high-risk" data processing.<sup>24</sup> These impact assessments require firms to evaluate the risks to an individual of the proposed data collection activities, the necessity of the collection, and mitigation measures to reduce risk.<sup>25</sup> Reflecting the GDPR's prioritization of individual data privacy rights above personal or collective benefits, nowhere does the law require firms to also assess the benefits of the data collection activities, either to individuals or to society.

Closing the data divide will require policymakers to look beyond narrow data privacy concerns to consider the myriad benefits of data collection and use for both organizations and individuals. Policymakers should avoid demonizing data collection and perpetuating stigmas surrounding data-driven technologies, such as labeling it "surveillance technology."<sup>26</sup> Rather than design policies restricting data, policymakers should look for ways to expand its equitable collection and use.

---

### **Recommendation: Pass National Data Privacy Legislation That Balances Privacy and Data Use**

Congress should pass federal data privacy legislation that balances individuals' privacy rights with data innovation. Ideally, a federal privacy law would establish basic consumer data rights, preempt state laws, ensure reliable enforcement, streamline regulation, and minimize the impact on innovation.<sup>27</sup> One step in this direction is the American Data Privacy and Protection Act, a draft bipartisan bill released for public discussion in June 2022 that would achieve many of these goals.<sup>28</sup> Congress should pass a refined version of this bill rather than allow the patchwork of state data privacy laws to expand and further curtail the collection and use of data and thereby exacerbate the data divide.

### **Recommendation: Establish Standards and Best Practices for Privacy-Enhancing Technologies**

Privacy enhancing technologies (PETs) are tools that allow for data use while reducing data risks.<sup>29</sup> PETs are one solution to the central data divide dilemma of balancing an increase in data collection and use while protecting individual privacy. Examples of PETs include secure multiparty computation (i.e., performing analysis on data held by different entities), de-identification of personal data (i.e., using statistical methods to anonymize personally identifiable data), and homomorphic encryption (i.e., performing mathematical operations on encrypted data).<sup>30</sup> Advances in PETs, as well as greater adoption and acceptance of these technologies, can foster increased use of sensitive data in sectors such as health care and education. For example, increased use of PETs could allow medical researchers to accelerate data-driven drug development, creating lifesaving medicine. While some PETs, such as differential privacy, have gained greater recognition and acceptance among scholars and practitioners, others have not. The Census Bureau has partnered with the United Nations Privacy-Enhancing Technology Lab to pilot different privacy-enhancing techniques. Establishing standards and best practices for PETs, including de-identification, would encourage greater use by all federal agencies, not just the Census.<sup>31</sup>

### **Recommendation: Reform Federal Sectoral Privacy Laws**

Researchers need access to data, but sectoral privacy laws often limit data collection and sharing in fields such as education and health care.<sup>32</sup> These laws can raise data collection costs, forcing researchers to make trade-offs, such as looking at easier-to-reach populations for which it costs less to obtain consent to use their data. Policymakers should reform privacy laws such as the Health Insurance Portability and Accountability Act (HIPAA) to better enable data sharing for beneficial purposes, such as improving patient care and medical research.<sup>33</sup> In education, the Family Educational Rights and Privacy Act (FERPA) and Higher Education Act govern access to educational data.<sup>34</sup> Although FERPA only covers information that directly identifies students, most institutions operate

---

under an assumption that all educational data is FERPA record.<sup>35</sup> Likewise, Congress prohibited the Department of Education from implementing a student-level database on higher educational outcomes under the Higher Education Act.<sup>36</sup> In the future, sectoral privacy laws should not obstruct data, particularly in high-value cases.

### **Create More Data**

A robust and equitable data economy is one in which most individuals live in data-rich environments. To build these environments, policymakers need to invest in programs that increase the amount of data collected by both the public and private sectors. Such programs should aim to increase the types of data collection devices used by individuals and communities and ensure that such distribution reaches traditionally underserved groups that stand to benefit the most. Data can propel economic growth and improve quality of life, but only if enough people can create, collect, and use it.

#### **Recommendation: Invest in Smart Cities**

Policymakers should invest in smart cities to improve data collection. Better data about communities can empower residents and civic leaders to make better decisions about local issues, such as crime and traffic. Sensors can automate data collection that cannot be easily or sufficiently collected manually. For example, air quality monitoring stations use smart sensors to collect continuous data points about the particulate matter in the air in real time.<sup>37</sup> A cross-agency taskforce in the federal government should create a comprehensive national strategy for the Internet of Things (IoT) to ensure that local communities can take full advantage of the opportunities created by IoT-connected devices.<sup>38</sup> Federal agencies, such as the Department of Transportation and the Department of Energy, should also fund research networks to develop smart cities solutions for urban challenges relating to things such as transportation and climate.<sup>39</sup> City governments should also create pilot programs for various sensor-based data collection initiatives, such as gunshot detection systems and food quality monitoring.<sup>40</sup>

#### **Recommendation: Identify and Procure More Private-Sector Data**

Government demand for data can create a market that boosts the supply of data. Federal agencies should issue an annual request for information soliciting feedback on what private-sector datasets could add value to their respective missions. For example, the Food and Drug Administration should examine what private sector data is necessary for monitoring consumer goods inventory and supply chain status for the products for which it maintains oversight such as infant formula.<sup>41</sup> The Office of Management and Budget should direct federal agencies to report on their use of private sector data as part of future annual action plans for complying with the FDS.<sup>42</sup> Greater use of private-sector data is especially important for filling data gaps.



---

### **Recommendation: Create Opportunities to Share Data Across Organizations and Sectors**

Many organizations, including businesses, nonprofits, government agencies, and universities, do not share their data with others. Federal agencies should create data sharing partnerships to increase data sharing among different groups to address important problems within their mission, such as to better track the spread of infectious diseases, understand climate change, and detect shipments of counterfeit products. Federal agencies should experiment with different data sharing models, such as data consortiums and data trusts, to encourage participation. Data sharing partnerships use formalized agreements to break down silos and aggregate data to create more data available for use while protecting sensitive or confidential information. These types of agreements can empower communities to have greater control over how their third parties use their data.<sup>43</sup>

### **Recommendation: Make It Easy for Americans to Donate Their Data**

Policymakers should make it easy for Americans to voluntarily contribute their sensitive, personal data to third parties. For example, although many Americans are willing to contribute their health data for use in medical research, few options exist for this type of donation. The Department of Health and Human Services should require all certifiable electronic health record systems to give patients the option to donate their data to third-party medical research. Additionally, the Office of Science and Technology Policy (OSTP) should search for other areas beyond health care where donating sensitive personal data may be useful and identify what types of technical and legal mechanisms can facilitate the process. Allowing people to donate their data empowers individuals and communities to address the data divide.

### **Recommendation: Create Datasets for High-Value AI Use Cases**

The United States needs more data to build successful AI models. Policymakers should fund the creation of high-quality, application-specific datasets to accelerate the development of AI-enabled tools. Innovation in certain industries such as health care, education, and defense depend on robust and representative datasets. Likewise, representativeness in data is also critical to ensuring that data-driven services such as AI work for all individuals and communities. For example, the development of new drugs requires representative datasets to ensure the medicine correctly benefits all groups. Likewise, AI for autonomous defense systems need updated datasets to ensure optimization and accuracy. Federal agencies should work with industry actors and civil society to develop shared datasets that can be an important resource for training new AI models.

### **Enhance Access to Data**

Policymakers should ensure that consumers and businesses alike can access data and put it to productive use. Access to data concerns the

---

availability of data sources to the public, whether it be government data or data collected by private actors. Policymakers should enhance access to data by supporting data portability policies, increasing open government data, and providing access to confidential government data (with appropriate safeguards).

**Recommendation: Create More Sector-Specific Data Portability Policies**

U.S. policymakers should create more sector-specific data portability policies at the national level, building off existing sectoral data protection laws that include data portability provisions. Data portability requires data controllers to make user data available in a standardized, machine-readable format.<sup>44</sup> Data controllers must also make user data available to consumers without any technical or legal restrictions on its use. One way to enable data portability is with open application programming interfaces (APIs). Open APIs allow third parties to access data on behalf of users from computer systems in a machine-readable format.<sup>45</sup> When practical, regulators should require the use of open APIs to implement data portability requirements.

Some sectors already have data portability requirements. For example, in the health care sector, the Health Information Technology for Economic and Clinical Health (HITECH) Act gives consumers the right to request their personal health information in a machine-readable electronic format and send that information to a designated third party.<sup>46</sup> Likewise, in the financial services sector, the Dodd-Frank Act requires that financial institutions provide consumers access to their financial records in a usable electronic form, although the Consumer Financial Protection Bureau (CFPB) still needs to do more rulemaking on the implementation of this requirement.<sup>47</sup> Policymakers should expand data portability in other sectors, such as the energy sector. For example, consumers should be able to access their utility data collected by smart meters in their homes to manage their energy consumption more effectively and thus save money.

**Recommendation: Support Open Data at State and Local Levels**

“Open data” refers to data made freely available for use without restrictions and plays a critical role in government transparency.<sup>48</sup> Open data benefits many actors, including businesses, civil society organizations, academics, and journalists. While federal agencies have made government data freely available and accessible by default, state and local governments should also continue developing and implementing open data policies.<sup>49</sup> Only 42 states have open data portals, and just 48 cities or counties across the United States publish open data.<sup>50</sup> The remaining states should develop open data policies and portals to make more data publicly available by default. Congress should create grant programs to help more cities launch open data initiatives.

---

### **Recommendation: Provide Access to Confidential Government Data With Appropriate Safeguards**

Although confidential government information should not be made publicly available, secure access to certain sensitive data should be made available to researchers. Only looking at aggregate data can hide important nuances between groups. As a result, policymakers should allow greater sharing of confidential government data with appropriate safeguards and expand existing examples of trusted data sharing to other federal agencies. For example, the Census allows the sharing of confidential microdata for use in social sciences research.<sup>51</sup> At present, researchers can access the data through secure Federal Statistical Research Data Centers. Under the 2018 Foundations for Evidence-based Policymaking Act, all federal statistical agencies must make their restricted data available after a standard application process.<sup>52</sup> Congress should monitor the implementation of the standard application process and ensure all covered agencies meet this requirement in a timely manner.<sup>53</sup>

Similarly, under the 2022 CHIPS and Science Act, the National Science Foundation must create a National Secure Data Service (NSDS) demonstration project to streamline data sharing across the federal government while retaining strong privacy and confidentiality practices.<sup>54</sup> Depending on the results of the demonstration's probationary period, policymakers should continue funding an NSDS to reduce the burden of data sharing while still protecting privacy. An NSDS operates under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2018 and the Privacy Act of 1974.<sup>55</sup> Allowing this type of formalized data sharing to continue means data that cannot be made public for privacy reasons can still be available to researchers under certain conditions.

### **Improve Data Quality**

Improving data quality—that is, ensuring data is fit for use, including accuracy, timeliness, precision, and representation—is important for closing the data divide and ensuring that all communities are included in important datasets. Some communities are left out from critical datasets, or the quality of collected data makes it less useful. In both instances, data quality inhibits certain individuals and groups from receiving the benefits of data and thus exacerbates the data divide.

### **Recommendation: Promote Data Interoperability Across Federal, State, and Local Governments**

Government statistics are often improperly formatted, incomplete, or inconsistent.<sup>56</sup> Policymakers should fix unnecessary variations in methodology and terminology across both the federal government and state and local governments that receive federal funding. For example, federal agencies use 23 variations to describe instances of sexual assault, with the differentiation often serving no substantial purpose.<sup>57</sup> These types

---

of terminology inconsistencies can also affect education data and obstruct effective decision-making by families and educators. States receive federal funding to build statewide longitudinal data systems that standardize education data within a state, but policymakers should require that these individual state systems also operate with the same lexicon and formatting to allow for cross-state interoperability and analysis. Policymakers should apply this type of standardization to other industries, such as labor and manufacturing, to create better data coverage for areas of national concern. Federal, state, and local governments should also pursue partnerships to promote data standards and data interoperability around common interests, including in such areas as transportation, education, the environment, and criminal justice.

#### **Recommendation: Strengthen Public-Private Data Standardization Practices**

Policymakers should also work to strengthen data standardization practices both between agencies and within the private sector. To do this, OSTP should convene a working group with key industry stakeholders to create a series of standards for critical sectors that both public and private organizations can use to make data more interoperable and useful. For example, the Digital Accountability and Transparency (DATA) Act of 2014 creates standards concerning the reporting of expenditures for government agencies so that agencies disclose spending in a streamlined manner.<sup>58</sup> OSTP should create a public-private working group to create consistent definitions and standard naming conventions for various topics in order to facilitate more communication between data systems and greater data sharing overall.

#### **Recommendation: Pass Demographic-Specific Data Protection Measures**

Many groups remain persistently underrepresented in data due to insufficient sampling or the exclusion of certain details about a population.<sup>59</sup> Policymakers should pass demographic-specific protections to formalize the representation of certain groups in data to fill historical data gaps. Representation in datasets impacts how well a service functions for or impacts a given community.

Some policymakers have begun introducing this type of legislation. For example, the LGBTQI+ Data Inclusion Act (H.R. 4176) passed the House of Representatives in July 2022 to ameliorate deficiencies in federal surveys relating to the collection of sexual orientation and gender identity information.<sup>60</sup> Likewise, the Equitable Data Working Group recommendations directed executive agencies to develop the Federal Evidence Agenda on LGBTQI+ Equity.<sup>61</sup> Higher-quality data for this group will help monitor population trends and identify community needs, such as LGBTQ-specific health care services. Likewise, bills introduced such as the Equitable Data Collection and Disclosure on COVID-19 Act of 2021 (H.R. 1370) sought to expand the collection of demographic information during

---

the pandemic in order to reduce disparities in health outcomes.<sup>62</sup> Congress should pass this type of legislation to ensure better representation in data.

#### **Recommendation: Use Synthetic Data to Fill Critical Data Gaps**

Investing in the use of synthetic data for sensitive datasets could help agencies fill critical data gaps, while protecting individual's privacy. Synthetic data refers to artificially created data that reflects real-world information.<sup>63</sup> This data has the same statistical properties as its real-world foil without containing any of the real information, meaning it can still be disaggregated or combined like a standard dataset while preserving privacy.<sup>64</sup> The Census Bureau already offers multiple products that use synthetic data, such as the OnTheMap application that maps where workers in the United States live.<sup>65</sup> The private sector also uses synthetic data when handling sensitive information. For example, privacy laws inhibit software companies from accessing patient data, so organizations building tools for electronic health record systems use stand-in data to test their products, and often face unforeseen difficulties when deploying it with actual patient data.<sup>66</sup> Synthetic data eliminates those difficulties as a high-quality stand in with the same statistical properties as the original data. Government chief technology and statistical officers should invest in synthetic data for the development of new products and services.

#### **Recommendation: Routinely Maintain Government Datasets**

Federal agencies should provide routine maintenance to datasets to increase the quality and usefulness of data. Just as the government treats the maintenance of infrastructure as an ongoing process that requires routine upkeep to ensure quality and safety, agency data collectors should continually monitor datasets to ensure they are updated, cleaned, and secured.<sup>67</sup> Moreover, agencies should maintain a list of known limitations and errors in datasets to clarify data quality and assist researchers seeking to use public data. Disclosing limitations of datasets will help those who use them understand who is not represented in the data, as well as how to improve them over time.

## **CONCLUSION**

Placing too much emphasis on only one solution area at a time will lead to an imbalance. Increasing the amount of data collected without improving data quality will create needless surveillance, and enhancing access to data on its own will be fruitless without better, more representative data. Moreover, closing the data divide will not be possible without rethinking and reframing privacy altogether. As explained, policymakers have largely found a successful approach to closing the digital divide, but have lacked a concerted, cohesive approach to closing the data divide. It's time to take the lessons learned from the digital divide and apply them to the data economy and ensure that no individuals and communities are left behind.

---

## APPENDIX

### Increase Digital Connectivity

Digital connectivity programs aim to maximize the number of households in the United States that can use broadband services. Connectivity policies account for the majority of funding aimed at closing the digital divide, with more than \$43.6 billion available for programs with connectivity as their exclusive purpose.<sup>68</sup>

Digital connectivity programs focus on broadband infrastructure planning and deployment, the speed of services, and the affordability of such services. Planning programs assess the feasibility of broadband deployment projects and preliminary engineering work.<sup>69</sup> Funding for deployment focuses on the construction of infrastructure and ongoing operations, which can include everything from network management to customer service and administration.<sup>70</sup>

Forty-two programs with broadband as their main purpose or one of their core purposes fall under the connectivity umbrella, including 11 ongoing programs and 5 created by recent legislation with broadband as their main purpose.<sup>71</sup> The FCC and the U.S. Department of Agriculture (USDA), among other agencies, run ongoing programs. For newly created programs, NTIA, the FCC, and numerous smaller agencies will be responsible for implementation. Connectivity programs can be nationwide or region-specific, or have eligible recipients be members of specific demographic groups, such as Native Americans or Native Hawaiians.

Notable ongoing connectivity programs include the FCC's High Cost Program and USDA's ReConnect Program. The High Cost program aims to subsidize telecommunications carriers for the cost of constructing, operating, and maintaining infrastructure in hard-to-reach areas and received \$28.3 billion in funding between FY 2015 and FY 2020.<sup>72</sup> The ReConnect Program focuses on smaller entities with a similar goal of constructing and operating broadband facilities for rural areas and received \$1.4 billion over the same period.<sup>73</sup> Newer programs have vastly increased the funding available for connectivity projects, with the Broadband Equity, Access, and Deployment Program directed by the NTIA receiving \$42.45 billion to support projects concerning planning, deployment, mapping, and adoption of broadband nationwide.<sup>74</sup> Likewise, the Tribal Broadband Connectivity Program received \$1 billion from the Consolidated Appropriations Act of 2021 and \$2 billion from the IJA in order to expand the adoption and deployment of broadband on tribal lands, among other endeavors.<sup>75</sup>

### Improve Digital Literacy

Digital literacy refers to the skills and ability to use information and communication technologies to find, evaluate, and communicate

---

information.<sup>76</sup> In the years since the National Broadband Plan, policymakers and other experts have come to recognize the importance of digital literacy in closing the digital divide. Although the 2010 plan outlines the importance of digital literacy skills in adopting broadband, gaps in digital literacy still remain, especially among certain age and income brackets.<sup>77</sup>

Five of the new programs created by recent legislation with broadband as their main purpose specifically target digital skills. For example, the Digital Equity Act provides \$2.75 billion to three grant programs to support the planning and implementation of digital equity plans to include more communities and spur greater adoption of broadband.<sup>78</sup> The act reveals that some communities will be left behind, barring action toward closing the digital literacy and skills gap. NTIA, the Department of Education, and smaller groups, such as the Institute of Museum and Library Services, govern most digital literacy programs.

### **Expand Access to Digital Devices**

In order to use broadband services, Americans need access to Internet-connected devices that can create, generate, and send information. Such devices include smartphones, tablets, and personal computers. A number of federal programs have sought to increase ownership of digital devices in underserved communities. This endeavor has become especially important in the wake of the COVID-19 pandemic that requires many students and professionals to complete their daily activities remotely. And yet, digital device use among adults in differing income brackets still exists. According to Pew Research Center, 24 percent of adults with annual incomes of less than \$30,000 report not owning a smartphone, and 41 percent lack a desktop or laptop computer.<sup>79</sup> In contrast, 97 percent of adults earning more than \$100,000 annually own a smartphone, and 92 percent own a desktop or laptop computer.

The influx of new funding for broadband programs emphasizes digital devices. Of programs with broadband as their main purpose, nine fall under the category of devices. For programs with broadband as one possible purpose, 25 relate to devices. Many connectivity programs include the need for digital devices, such as NTIA's Tribal Broadband Connectivity Program. Programs run by the FCC, such as the COVID-19 Telehealth Program, Affordable Connectivity Program, Emergency Connectivity Fund, and Connected Care Pilot, all focus on affordability and devices in order to ensure individuals and communities can access essential health care services via broadband.

---

## REFERENCES

1. Mary Madden and Lee Rainie, “Americans’ Views About Data Collection and Security” (Pew Research Center, May 2015), <https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/>.
2. Government Accountability Office (GAO), *National Strategy Needed to Guide Federal Efforts to Reduce Digital Divide* (Washington DC: GAO, May 2022), <https://www.gao.gov/products/gao-22-104611>.
3. “The Clinton-Gore Administration: From Digital Divide to Digital Opportunity,” Clinton Whitehouse Archives, accessed March 5, 2023, <https://clintonwhitehouse4.archives.gov/WH/New/digitaldivide/> (site discontinued).
4. Federal Communications Commission (FCC), *Connecting America: The National Broadband Plan* (Washington DC: FCC, March 2010), <https://transition.fcc.gov/national-broadband-plan/national-broadband-plan.pdf>.
5. Ibid.
6. Richard Bennett, Luke A. Stewart, and Robert D. Atkinson, “The Whole Picture: Where America’s Broadband Networks Really Stand” (ITIF, February 2013), <https://itif.org/publications/2013/02/12/whole-picture-where-americas-broadband-networks-really-stand/>.
7. Jessica Dine and Joe Kane, “The State of US Broadband in 2022: Reassessing the Whole Picture” (ITIF, December 2022), <https://itif.org/publications/2022/12/05/state-of-us-broadband-in-2022-reassessing-the-whole-picture/>.
8. National Urban League, “The Lewis Latimer Plan” (National Urban League, January 2021), <https://nul.org/program/lewis-latimer-plan>.
9. “Summary of Broadband Service Provider Funding Opportunities,” Bradley, accessed February 2023, <https://www.bradley.com/-/media/files/insights/publications/2021/11/iija-broadband-summary.pdf>.
10. GAO, *National Strategy Needed to Guide Federal Efforts to Reduce Digital Divide*.
11. Ibid.
12. Madeline Hughes, “New FCC Programs Tackle Affordability Gap in Digital Divide,” *The Well News*, August 5, 2022, <https://www.thewellnews.com/technology/new-fcc-programs-tackle-affordability-gap-in-digital-divide/>.
13. Daniel Castro, “The Rise of Data Poverty in America” (Center for Data Innovation, September 2014), <https://www2.datainnovation.org/2014-data-poverty.pdf>.
14. Gillian Diebold, “Closing the Data Divide for a More Equitable U.S. Digital Economy” (Center for Data Innovation, August 2022),



- 
- <https://datainnovation.org/2022/08/closing-the-data-divide-for-a-more-equitable-u-s-digital-economy/>.
15. “Federal Data Strategy,” accessed February 2023, <https://strategy.data.gov/>.
  16. Diebold, “Closing the Data Divide for a More Equitable U.S. Digital Economy.”
  17. Morgan Stevens, “Better Data on College Graduates’ Financial Outcomes Can Help Future Students Make Better Decisions,” Center for Data Innovation, October 25, 2022, <https://datainnovation.org/2022/10/better-data-on-college-graduates-financial-outcomes-can-help-future-students-make-better-decisions/>.
  18. Laura Palk and Krishnamurty Muralidhar, “A Free Ride: Data Brokers’ Rent-Seeking Behavior and the Future of Data Inequality,” *Vanderbilt Journal of Entertainment & Technology Law*, Vol. 20, No. 3, 2018, <https://scholarship.law.vanderbilt.edu/jetlaw/vol20/iss3/4>.
  19. Biometric Information Privacy Act, 740 ILCS, 2008, <https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57>.
  20. Lauren Harrison, “When Do Privacy Regulations Go Too Far?” Government Technology, n.d., <https://www.govtech.com/opinion/when-do-privacy-regulations-go-too-far.html>.
  21. “Enterprise Income Verification (EIV) System” (Notice H 2013-06 from U.S. Department of Housing and Urban Development, March 2013), <https://www.hud.gov/sites/documents/13-06HSGN.PDF>; Michael Turner and Patrick Walker, “Potential Impacts of Credit Reporting Public Housing Rental Payment Data” (Washington, D.C., U.S. Department of Housing and Urban Development, October 2019), <https://www.huduser.gov/portal/publications/Potential-Impacts-of-CreditReporting.html>.
  22. “What is GDPR, the EU’s new data protection law?” (GDPR), accessed February 2023, <https://gdpr.eu/what-is-gdpr/>.
  23. “Data Protection” (EDPS), accessed January 2023, [https://edps.europa.eu/data-protection/data-protection/glossary/d\\_en](https://edps.europa.eu/data-protection/data-protection/glossary/d_en).
  24. “Privacy Impact Assessment” (GDPR), accessed January 2023, <https://gdpr-info.eu/issues/privacy-impact-assessment/>.
  25. Art. 35 GDPR: “Data protection impact assessment,” intersoft consulting, <https://gdpr-info.eu/art-35-gdpr/>
  26. Commercial Surveillance and Data Security Rulemaking, 87 Fed. Reg. 51273 (proposed August 22, 2022), <https://www.federalregister.gov/documents/2022/08/22/2022-17752/trade-regulation-rule-on-commercial-surveillance-and-data-security>.
  27. Daniel Castro, Testimony to the House Administration Committee on “Big Data: Privacy Risks and Needed Reforms in the Public and Private Sectors,”

- 
- February 16, 2022, <https://itif.org/publications/2022/02/16/testimony-house-administration-committee-big-data-privacy-risks-and-needed/>.
28. American Data Privacy and Protection Act, H.R. 8152, 117th Congress, (2022) <https://www.congress.gov/bill/117th-congress/house-bill/8152/text>.
  29. The Royal Society, “Protecting Privacy in Practice” (Royal Society, March 2019), <https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Protecting-privacy-in-practice.pdf>.
  30. Office of Science and Technology Policy (OSTP), “National Strategy to Advance Privacy -Preserving Data Sharing and Analytics” (Washington D.C.: OSTP, March 2023), <https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Strategy-to-Advance-Privacy-Preserving-Data-Sharing-and-Analytics.pdf>.
  31. Simson Garfinkel, “De-Identification of Personal Information” (Washington D.C.: National Institute for Standards and Technology, October 2015), <https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>.
  32. Roberta Ness, “Influence of the HIPAA Privacy Rule on Health Research,” *JAMA*, (2007), 2164–2170, <https://jamanetwork.com/journals/jama/fullarticle/209447>.
  33. Donald Berwick and Martha Gaines, “How HIPAA Harms Care, and How to Stop It,” *JAMA* (2018) <https://jamanetwork.com/journals/jama/article-abstract/2686002>.
  34. Family Educational Rights and Privacy Act (FERPA), 1974, <https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>.
  35. Ibid.
  36. Stevens, “Better Data on College Graduates’ Financial Outcomes Can Help Future Students Make Better Decisions.”
  37. “Basic Information about Air Emissions Monitoring,” Environmental Protection Agency, last modified August 11, 2022, <https://www.epa.gov/air-emissions-monitoring-knowledge-base/basic-information-about-air-emissions-monitoring>.
  38. Joshua New and Daniel Castro, “Why Countries Need National Strategies for the Internet of Things” (Center for Data Innovation, December 2015), <https://www2.datainnovation.org/2015-national-iot-strategies.pdf>.
  39. Colin Cunliff, Ashely Johnson, and Hodan Omaar, “How Congress and the Biden Administration Could Jumpstart Smart Cities with AI” (ITIF, March 2021), <https://www2.itif.org/2021-smart-cities-ai.pdf>
  40. Qiannan Zhang et al., “A Case Study of Sensor Data Collection and Analysis in Smart City: Provenance in Smart Food Supply Chain,” *Sage Journals*, November 11, 2013, <https://journals.sagepub.com/doi/10.1155/2013/382132>.
  41. Gillian Diebold, “Better Data Sharing Could Help Predict and Prevent National Shortages in Baby Formula and Other Critical Consumer Goods,” *Center for Data Innovation*, August 15, 2022, <https://datainnovation.org/2022/08/better-data-sharing-could-help->

- 
- predict-and-prevent-national-shortages-in-baby-formula-and-other-critical-consumer-goods/.
42. “Federal Data Strategy,” accessed February 2023, <https://strategy.data.gov/>.
  43. “Advancing an Equitable Government,” performance.gov, accessed April 2023, <https://www.performance.gov/equity/#action-plans>.
  44. Daniel Castro, “Improving Consumer Welfare with Data Portability” (Center for Data Innovation, November 2021), <https://www2.datainnovation.org/2021-data-portability.pdf>.
  45. Daniel Castro and Michael Steinberg, “Blocked: Why Some Companies Restrict Data Access to Reduce Competition and How Open APIs Can Help” (Center for Data Innovation, November 2017), <https://www2.datainnovation.org/2017-open-apis.pdf>.
  46. “Health Information Technology for Economic and Clinical Health Act,” Pub. L. No. 111-5, 123 Stat. 226 (2009), <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveridentities/hitechact.pdf>.
  47. “Open Banking, Data Sharing, and the CFPB’s 1033 Rulemaking,” Congressional Research Service, September 9, 2021, <https://crsreports.congress.gov/product/pdf/IN/IN11745>.
  48. “What is Open Data?” Open Data Handbook, accessed January 2023, <https://opendatahandbook.org/guide/en/what-is-open-data/>.
  49. Foundations for Evidence-Based Policymaking Act, H.R. 4174, 115th Congress (2018), <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
  50. “Open Government,” Data.gov, accessed January 2023, <https://data.gov/open-gov/>.
  51. “Restricted-Use Data,” Census Bureau, accessed March 2023, <https://www.census.gov/topics/research/guidance/restricted-use-microdata.html>.
  52. Foundations for Evidence-Based Policymaking Act, H.R. 4174, 115th Congress (2018), <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
  53. “The Standard Application Process,” National Center for Science and Engineering Statistics, accessed April 2023, <https://nces.nsf.gov/about/standard-application-process>.
  54. Jory Heckman, “National Secure Data Service comes into focus through semiconductor bill,” *Federal News Network*, September 9, 2022, <https://federalnewsnetwork.com/federal-insights/2022/09/national-secure-data-service-comes-into-focus-through-semiconductor-bill/>.
  55. Data Foundation, “Congress Authorizes Establishment of National Secure Data Service to Improve Data Analytics,” news release, July 28, 2022, <https://www.datafoundation.org/press-releases/congress-authorizes-establishment-of-national-secure-data-service-to-improve-data-analytics/2022>.
  56. “Briefing Paper on Open Data and Data Quality” (The Center for Open Data Enterprise, 2016)

- 
- <http://reports.opendataenterprise.org/BriefingPaperonOpenDataandImprovingDataQuality.pdf>.
57. Joshua New, “How Common is Sexual Assault in the United States? The Answer Depends on Who You Ask” (Center for Data Innovation, September 2016), <https://datainnovation.org/2016/09/how-common-is-sexual-assault-in-the-united-states-the-answer-depends-on-who-you-ask/>.
  58. DATA Act, S.994, 113th Congress (2013), <https://www.congress.gov/bill/113th-congress/senate-bill/994>.
  59. Diebold, “Closing the Data Divide for a More Equitable U.S. Digital Economy.”
  60. LGBTQI+ Data Inclusion Act, H.R. 4176, 117th Congress (2021), <https://www.congress.gov/bill/117th-congress/house-bill/4176/text>.
  61. National Science and Technology Council Subcommittee on Equitable Data, “Progress on Implementation of the Recommendations of the Equitable Data Working Group” (Washington DC: NSTC, March 2023), <https://www.whitehouse.gov/wp-content/uploads/2023/03/Progress-on-Equitable-Data-Mar2023.pdf>.
  62. Equitable Data Collection and Disclosure on COVID-19 Act, H.R. 1370, 117th Congress (2021), <https://www.congress.gov/bill/117th-congress/house-bill/1370>.
  63. Gerard Andrews, “What is Synthetic Data?,” *Nvidia*, June 8, 2021, <https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/>.
  64. “What are Synthetic Data?” Census Bureau, last modified May 27, 2021, <https://www.census.gov/about/what/synthetic-data.html>.
  65. “OnTheMap,” Census Bureau, accessed February 2023, <https://onthemap.ces.census.gov/>.
  66. MIT Laboratory for Information and Decision Systems, “The real promise of synthetic data,” news release, October 16, 2020, <https://news.mit.edu/2020/real-promise-synthetic-data-1016>.
  67. National Academies of Sciences, Engineering, and Medicine, *Toward a 21st Century National Data Infrastructure: Mobilizing Information for the Common Good* (Washington, DC: The National Academies Press, 2022), <https://www.nationalacademies.org/our-work/toward-a-vision-for-a-new-data-infrastructure-for-federalstatistics-and-social-and-economic-research-in-the-21st-century>.
  68. Government Accountability Office, *National Strategy Needed to Guide Federal Efforts to Reduce Digital Divide* (Washington DC: GAO, May 2022), <https://www.gao.gov/products/gao-22-104611>.
  69. Ibid.
  70. Ibid.
  71. Ibid.
  72. “Universal Service for High Cost Areas – Connect America Fund,” Federal Communications Commission, last modified April 4, 2023,

- 
- <https://www.fcc.gov/general/universal-service-high-cost-areas-connect-america-fund>.
73. “ReConnect Loan and Grant Program,” U.S. Department of Agriculture, accessed January 2023, <https://www.usda.gov/reconnect>.
  74. “Broadband Equity, Access, and Deployment (BEAD) Program,” Broadband USA, accessed December 2022, <https://broadbandusa.ntia.doc.gov/resources/grant-programs/broadband-equity-access-and-deployment-bead-program>.
  75. “Tribal Broadband Connectivity Program,” Broadband USA, accessed December 2022, <https://broadbandusa.ntia.doc.gov/resources/grant-programs/tribal-broadband-connectivity-program>.
  76. “Definitions,” National Digital Inclusion Alliance, accessed December 2022, <https://www.digitalinclusion.org/definitions/>.
  77. Emily Vogels and Monica Anderson, “Americans and Digital Knowledge” (Pew Research Center, October 2019), <https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/> and John Horrigan, “Digital Readiness Gaps” (Pew Research Center, September 2016), <https://www.pewresearch.org/internet/2016/09/20/digital-readiness-gaps/>.
  78. “Digital Equity Act Programs,” Broadband USA, accessed January 2023, <https://broadbandusa.ntia.doc.gov/resources/grant-programs/digital-equity-programs>.
  79. Emily Vogels, “Digital divide persists even as Americans with lower incomes make gains in tech adoption,” *Pew Research Center*, June 22, 2021, <https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/>.

---

## ABOUT THE AUTHORS

Gillian Diebold is a Policy Analyst at the Center for Data Innovation. She holds a B.A. from the University of Pennsylvania, where she studied Communication and Political Science.

Daniel Castro is the director of the Center for Data Innovation and vice president of the Information Technology and Innovation Foundation. He has a B.S. in foreign service from Georgetown University and an M.S. in information security technology and management from Carnegie Mellon University.

## ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation studies the intersection of data, technology, and public policy. With staff in Washington, London, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, artificial intelligence, and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

**Contact: [info@datainnovation.org](mailto:info@datainnovation.org)**

**[datainnovation.org](http://datainnovation.org)**