

Leveraging Demographic Data

By Joe Paul, Natalie Rogers and Morgan Thomas

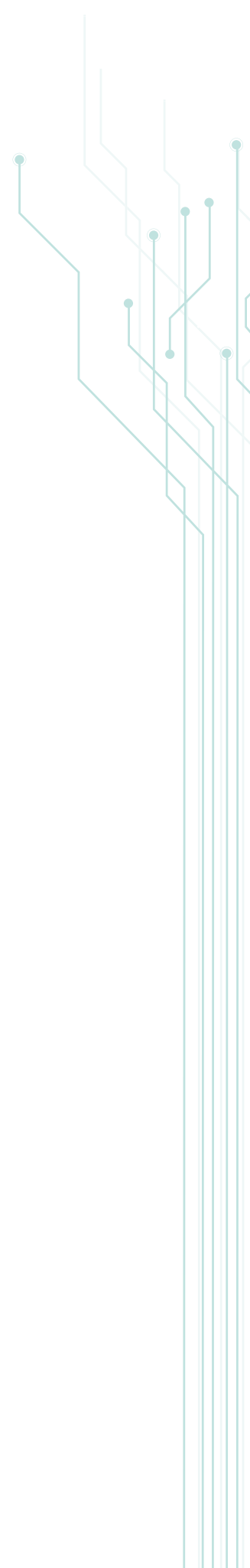
1. Introduction

At the center of data quality is the need for an understanding of the vast amount of data collected by government agencies and how that data is verified and used by decision-makers for resource allocation and policy improvements. A key type of data referenced to make these decisions is demographic data. Briefly, demographic data refers to statistical information about populations or groups of people, including various characteristics such as age, gender, ethnicity, income, education level, marital status and more. This data provides insight into the programs, services and activities individuals are engaged in and the characteristics of that population.

Demographic data serves as a foundation for informed decision-making across various sectors, guiding policies, programs and strategies tailored to meet the diverse needs of populations. The role that demographics play in the allocation of resources is significant. Thus, the importance of good demographic data collection cannot be understated. Data collection is crucial for understanding and addressing the needs, behaviors and preferences of different demographic groups. It provides valuable insights for decision-making in various sectors, including policymaking, resource allocation, marketing and social services.

The [Office of Management and Budget's Statistical Policy Directive No. 15 \(SPD15\)](#) serves as a foundational framework for ensuring the quality, integrity and comparability of federal statistical data across government agencies. The release of the Office of Management and Budget's revisions to SPD15 in March 2024 marks a great opportunity to reset data collection standards and incorporate updated measures and practices. Established in 1980, SPD15 outlines standardized principles and practices for the collection, analysis and dissemination of statistical information, aiming to uphold the credibility and reliability of statistical data used for policymaking, research and decision-making.

By providing clear guidelines for statistical agencies, SPD15 promotes consistency and coherence in statistical programs, facilitating accurate measurement and meaningful interpretation of key economic, social and demographic indicators by providing program oversight, regarding methods, definitions, classifications and reporting standards. Furthermore, the directive emphasizes the importance of protecting the confidentiality and privacy of individual respondents ensuring that sensitive information is safeguarded while still allowing for the robust analysis of aggregate data.



Compliance with SPD15 is essential for maintaining public trust in government statistics as it demonstrates a commitment to transparency and professional standards in the production and dissemination of statistical information. Moreover, adherence to the directive enhances the effectiveness of data-driven decision-making, as policymakers, researchers and the public can have confidence in the accuracy, relevance and reliability of federal statistical data. As such, SPD15 plays a critical role in supporting evidence-based policymaking, fostering accountability and promoting informed public discourse on issues of national importance. Starting March 2024, agencies have five years to complete their compliance with the revised directive.

This white paper underscores the significance of collecting demographic data and explores effective methods for its collection. It will delve into the role of demographic data in decision-making processes, highlight its benefits, address challenges and outline best practices for ethical data collection. We will demonstrate to leadership at the state and federal levels and other stakeholders of Registered Apprenticeship Programs (RAP) the necessity of sound demographic data collection to determine program effectiveness as well as identifying any deficiencies in serving targeted populations.

2. Understanding Demographic Data

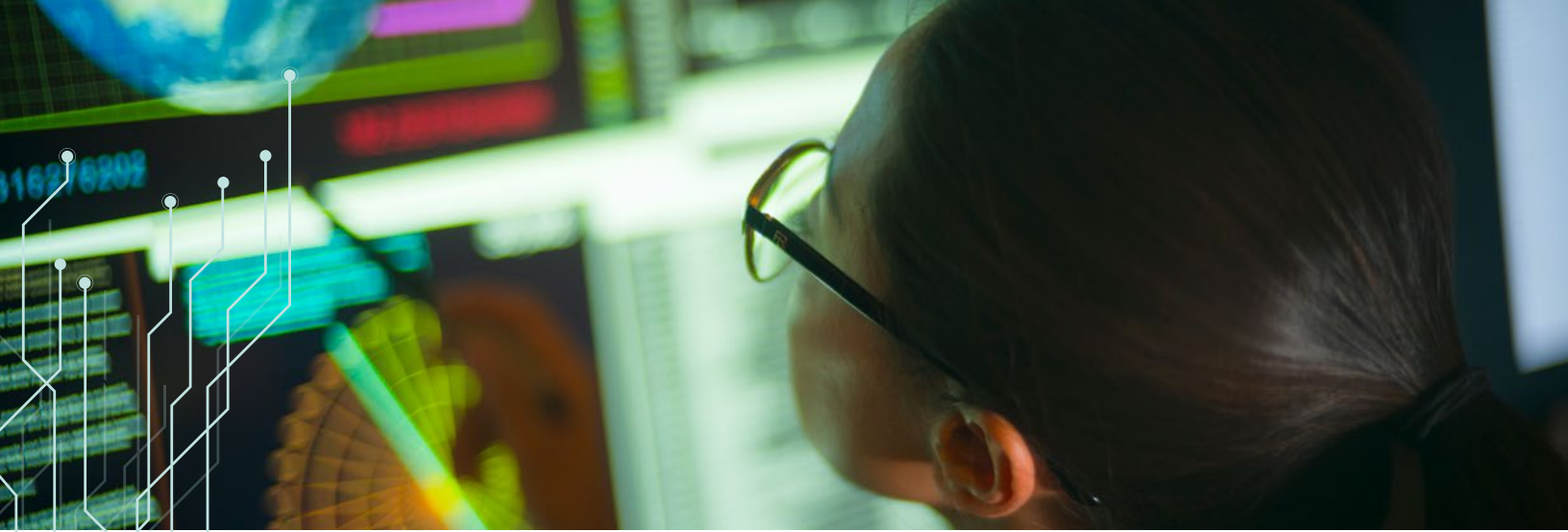
Demographic data refers to statistical information about a population, often categorized by various characteristics. These characteristics include age, gender, race, ethnicity, education level, income, occupation, marital status, household size and geographic location, in addition to more nuanced factors like disability status, immigration status and language spoken at home. The collection and analysis of demographic data provides valuable insights into the composition and dynamics of a population, helping to inform policies, programs and business decisions.

Demographic data can be further broken down into two main types: static and dynamic. Static demographic data represents characteristics of a population at a specific point in time, such as the population size, age distribution and racial composition. Dynamic demographic data captures changes in population characteristics over time, including birth rates, death rates, migration patterns and changes in socioeconomic status.

Monitoring inequalities is increasingly recognized as a key step in providing equitable access to government services, including employment programs like Registered Apprenticeships. Agencies use data collected by their programs to understand their clients and the effectiveness of their programs. There are various forms of demographic data, including:

- Sociodemographic data.
- Economic data.
- Health data.
- Cultural data.

Sociodemographic data encompasses factors related to both social and demographic characteristics. This includes age, gender, race, ethnicity, marital status, education level, occupation, income, and household size and composition. Sociodemographic data helps to understand the social and economic makeup of a population, including disparities and inequalities. For example, analyzing sociodemographic data can reveal patterns of poverty, educational attainment and employment opportunities within different demographic groups.



Economic data also encompasses factors related to both social and demographic characteristics. This can include age, gender, race, ethnicity, marital status, education level, occupation, income and household size. A lot of economic data is collected by the [Bureau of Labor Statistics](#) within the U.S. Department of Labor through the Current Employment Statistics survey of 300,000 businesses (non-farm). Together with the Occupational Employment and Wage Statistics, these surveys provide data to decision-makers on employment levels, earnings, average hours weekly hours worked and overall economic performance.

Health data pertains to information about the health status, behaviors and health care utilization patterns of a population. This includes indicators such as the prevalence of chronic diseases, vaccination rates, access to health care services, lifestyle factors (e.g., smoking, diet, exercise), mental health indicators and mortality rates. Health data is crucial for assessing health care needs, planning public health interventions and evaluating the effectiveness of health programs and policies. Additionally, health data can guide efforts to reduce health inequities by revealing disparities in health outcomes among different demographic groups.

Cultural data encompasses aspects of identity, beliefs, values, traditions and practices within a population. This includes language spoken, religion, cultural heritage, immigration status, acculturation levels and participation in cultural activities. Cultural data helps to understand the diversity and cultural dynamics of a population, guiding efforts to promote cultural competence, inclusion and diversity in various sectors such as education, health care and social services. Understanding cultural data is essential for developing culturally sensitive policies and programs that respect the needs and preferences of diverse communities.

For governments, demographic data is used to allocate funding, plan infrastructure development and design social welfare programs. For example, census data is used to determine the distribution of seats in the legislative branch and allocate federal funding to states and localities. In business, demographic data helps companies identify market segments, tailor products and services to specific customer needs, and expand into new markets. In health care, demographic data informs public health initiatives, guides resource allocation in health care systems and supports epidemiological research. In education, demographic data is used to assess student needs, plan curriculum and instruction and allocate resources to schools and districts.

Understanding demographic data is essential for making informed decisions that promote equity, efficiency and effectiveness in addressing the needs of diverse populations. By leveraging demographic data, decision-makers can develop evidence-based strategies that improve outcomes and enhance the well-being of individuals and communities. This is why collecting accurate demographic data is crucial to agencies and their decision-making processes.

3. Importance of Collecting Demographic Data

Collecting demographic data is critical and offers a range of benefits that can be leveraged to better serve diverse populations. Quality demographic data can allow for more informed policy decisions, more efficient resource allocation, tailored marketing strategies and a better understanding of community disparities.

Incorporating demographic data into policymaking ensures that policy decisions are evidence-based. It can be used to [help craft policy responses](#) to complex problems by allowing policymakers to better understand the scope of the issue and who is being impacted. In recent years, more states have been engaging in an [evidence-based policymaking](#) model that systemically incorporates findings from program evaluations and outcome analysis to guide policy and funding decisions. Leveraging demographic data is a key part of this process.

Policymakers and government agencies are often working with limited resources. Demographic data can provide useful insights to ensure that limited resources have the most impactful reach possible. Demographic data can also demonstrate who funding is going to and how it is impacting the population. Quality demographic data helps organizations better understand who they are serving and what their needs are. Having a strong understanding of the population an organization is attempting to reach is key to developing tailored outreach strategies that are the most effective at connecting with the intended audience.

Demographic data can also help demonstrate program and policy effectiveness. If collected over time, agencies can measure the impact of their choices and outcomes of their outreach efforts. For example, state apprenticeship agencies looking to recruit more people with disabilities into apprenticeship programs can track the number of apprentices who self-identify as a person with a disability over time to see if the numbers are increasing as they try different strategies.

Accurate demographic data offers opportunities to understand where disparities exist. Different groups of people often have varying levels of access to supportive services and can be impacted by policies disproportionately. Leveraging demographic data has been [especially impactful for health care organizations](#), which have organized available demographic data by patient race, ethnicity, language, age, sex, health literacy, sexual orientation, gender identity, socioeconomic status and geography to improve care and understand inequities. This data allows health care organizations to discover differences in care and outcomes across patient groups, plan equity-focused care transformations and measure the impact of their efforts.



4. Challenges in Collecting Demographic Data

Despite the importance of demographic data, many barriers can inhibit the collection and application of high-quality, accurate data. These challenges include privacy concerns from participants, inaccurate and unreliable data, ethical considerations, how identities are categorized, technological barriers and financial constraints.

PRIVACY CONCERNS

To collect demographic data, participants must be willing to disclose information about themselves. Research shows [that most Americans hold strong values](#) about their privacy and data security. They want to understand who is gathering information about them and what that information is going to be used for.

In some cases, participants fear volunteering certain demographic information about themselves because they believe that information might be used against them. An example of this is the self-identification of disabilities. Researchers found that [73% of respondents with disabilities](#) reported fear that disability self-identification would cause them to not get hired or to get fired. This fear might prevent someone from self-identifying as having a disability in demographic data collection. As a result, the data might not accurately reflect the population. To overcome these challenges, it is important to be transparent about why the data is being collected and how it will be used.

DATA COLLECTION SYSTEMS

Data systems can be old and inflexible. Outdated technologies can make it hard to access and interpret data. In some cases, even if a system has all the data needed, it might not be able to run new types of reports to analyze the data. When technology is out of date, resources are needed to implement new systems, which can be a time-consuming and expensive endeavor.

No data collection system is perfect. How data is collected, where it is stored and the resources used to analyze it can [pose challenges](#). Data can often be entered incorrectly or not at all. When there are mistakes or inaccuracies in data collection, it can be difficult to analyze the data properly, if at all.

Access to resources is another challenge for collecting and analyzing demographic data. Organizations need staff who have the knowledge and the capacity to properly use and apply data. If the resources aren't there to translate the data, it cannot be put to good use.

DISCLOSURE RELIABILITY

One of the biggest challenges in collecting demographic data is how to properly categorize different social identities, including race, ethnicity, sexual orientation and gender identity. These identities are complex and evolving. Oftentimes, demographic questions are written in a way [that ignores the complexity of identity](#). This can lead to frustration and confusion when responding to an individual who does not see their identity reflected in the question. This can also [decrease the number of participants willing to take the survey](#), decrease response rates to individual questions, and impact the accuracy of responses.

[Research has shown](#) that a growing number of people do not identify with historically used Office of Management and Budget (OMB) race and ethnicity categories. The 1997 OMB requirements for race and ethnicity included two groups for ethnicity and a minimum of five options for race. In recent years, [U.S. Census Bureau data](#) has revealed a significant increase in the number of people identifying as two or more races or some other race.



The OMB 1997 standards call for a two-question approach. In this format, respondents indicate which of five race categories (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White) they identify as, with the option to select multiple races. In a second question, respondents indicate whether they are of Hispanic or Latino ethnicity. Under these standards, people with Hispanic ethnicity and people from the Middle East and North Africa often selected other race because [there was not an available category they identified with](#).

In March 2024, OMB announced [revisions to its standards](#) for data collection on ethnicity and race in an effort to offer options that more accurately reflect the variety of racial and ethnic identities. The revisions included using a single combined question for race and ethnicity, adding Middle Eastern or North African as a category, clarifying instructions for individuals to select multiple racial and ethnic categories that represent their identity and requiring collection of more detail beyond the minimum categories.

ETHICAL CONSIDERATIONS FOR DEMOGRAPHIC DATA COLLECTION

Ethical collection of demographic data is crucial. Data collected about individual characteristics, including age, gender and ethnicity, must follow a process that respects the rights, privacy and autonomy of the person. Using the principles of [hermeneutic phenomenology](#), the person is as unique as their life story, researchers and their audiences should be made aware of the need to validate the person and their rights as well as understand how they identify themselves.

Researchers should take steps to minimize any potential harm or discomfort to participants. This may involve avoiding intrusive or sensitive questions, providing support resources for participants who may experience distress or ensuring that data collection methods are respectful and non-discriminatory.



5. Informed Consent and Transparency

Participants “[should be neither deceived nor coerced](#),” and fully informed about the purpose of data collection, how their data will be used, and any potential risks or benefits associated with participation. They should have the option to provide consent freely and without coercion. Researchers should be transparent about their data collection methods, including how data will be used, who will have access to it and any limitations or biases inherent in the process.

Anonymity and confidentiality measures should be in place to safeguard the privacy of participants’ data. This includes using secure data storage systems, anonymizing data whenever possible and limiting access to sensitive information to authorized personnel only.

DATA SECURITY MEASURES

Measures should be in place to protect data from unauthorized access, misuse, or theft. This includes encrypting data during transmission and storage, regularly updating security protocols and complying with relevant data protection regulations. The 2024 [X-Force Threat Intelligence Index](#) found “a 71% increase in cyber-attacks leveraging stolen identities in 2023 compared to 2022.”

REGULATORY COMPLIANCE (E.G., GDPR, HIPAA, FERPA)

Participants should have the right to control how their data is used and shared. Researchers should respect participants’ preferences regarding data sharing and take steps to ensure that they can provide input on how their data is used. Participation in data collection should be voluntary and individuals should have the right to withdraw from the study at any time without facing negative consequences. The data collection processes should be designed to ensure fairness and equity among participants. This includes avoiding bias in sampling methods, being inclusive of diverse populations, and considering the potential impact of data collection on marginalized or vulnerable groups.

By adhering to these ethical principles, agencies and researchers can ensure that demographic data collection processes are conducted in a manner that respects the rights of participants while still allowing for valuable insights to be gained from the data.

6. Effective Methods for Collecting Demographic Data

Effective demographic data collection requires a combination of traditional and innovative methods tailored to the specific context and objectives of the research. While traditional methods like surveys and censuses remain valuable, innovative approaches such as administrative data analysis, geospatial data utilization, social media analysis and machine learning offer opportunities to enhance data collection efficiency and accuracy. By leveraging a diverse set of data sources and methodologies, researchers can obtain a more comprehensive understanding of population demographics and inform evidence-based decision-making processes.

TRADITIONAL METHODS

The traditional methods of data collection are censuses and surveys. A census is a comprehensive method of collecting demographic data that aims to enumerate the entire population within a specified geographic area. Census data provides a snapshot of population characteristics at a particular

point in time and serves as a benchmark for various policy and planning decisions. However, conducting a census requires significant resources and may face challenges, such as undercounting certain population groups. The U.S. census occurs once every 10 years and is conducted by the U.S. Census Bureau. To expand the data that is collected between census periods, the U.S. Census Bureau also administers the American Community Survey.

Surveys can be administered through various channels, including face-to-face interviews, telephone interviews, and mail-in and online surveys. Surveys remain a common method for collecting demographic data that allow researchers to gather detailed demographic information, such as age, gender, education, income, ethnicity and household composition. However, survey response rates have been declining in recent years, posing a challenge to data collection efforts.

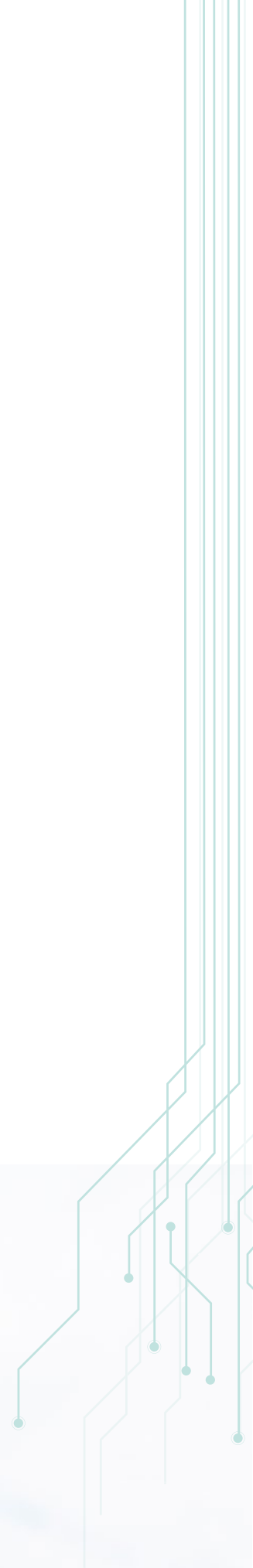
INNOVATIVE METHODS

Innovation plays a key role in collecting complete and accurate data sets. Through novel use of administrative data, new geospatial data techniques, social media data, new machine learning and AI, researchers can expand the collection methods of data.

Administrative data collected by government agencies and organizations for administrative purposes can serve as a valuable source of demographic information. Examples include tax records, birth and death certificates, school enrollment data and health care records. By linking and analyzing existing administrative data sources, researchers can obtain rich demographic insights while minimizing respondent burden and data collection costs. Participants enrolling in a Registered Apprenticeship Program are prime examples of collecting data for administrative purposes.

Geospatial data, which includes information about the geographic location of individuals and communities, can supplement traditional demographic data collection methods. Geospatial data sources, such as satellite imagery, GPS tracking and mobile phone data, offer opportunities to understand population distribution, mobility patterns and spatial disparities. However, privacy concerns and ethical considerations must be carefully addressed when using geospatial data.

Analysis of social media platforms generates vast amounts of user-generated data that can provide insights into demographic characteristics and behaviors. Analyzing social media data allows researchers to identify trends, sentiments and interactions among different population groups. However, biases in social media usage and data privacy issues need to be considered when interpreting findings from social media analysis.



Machine learning techniques can enhance the collection and analysis of demographic data by automating data processing tasks, identifying patterns and making predictions. Data integration approaches enable researchers to combine data from multiple sources, such as surveys, administrative records and social media, to create comprehensive demographic profiles. However, ensuring data quality, privacy protection and algorithm transparency are important considerations in employing machine learning for demographic data collection.

7. Improving Collection Efforts

We will now look at examples of how some organizations have used data to improve their service delivery. A tried and tested way to improve an organization's understanding of where they stand in using data to improve services and data collection is to look at examples of entities that have achieved success in using data. There are [several methods](#) used in both the public and private sectors, including identification of pain points and gaps in service delivery; implementation of a customer feedback system and service quality metrics; development of an empowered and well-trained staff; innovation and leveraging technology; and creation of a culture of service excellence.

In the private sector, we'll look at several companies that score high in customer satisfaction and customer loyalty. These companies have learned to leverage data to understand their customers and provide innovative solutions and maintain high levels of customer engagement. While not all these examples can be directly applied to public sector solutions, we can glean behaviors and methods.

CASE STUDY

In their article on ["Improving Patient Race and Ethnicity Data Capture to Address Health Disparities,"](#) Vega Perez, et al., outlined an approach that saw a 76% improvement in the completion of demographic data fields. The researchers looked at their understanding of why disparities by race and ethnicity still exist despite attempts to promote health equity. They identified the need for the "establishment of the basic information needed to identify and measure those differences, i.e., the accurate capture of race and ethnicity information of all patients." They also pointed out that their study incorporated scalability in their design while being cognizant of the reliability of the data.

"The project consisted of five phases outlining a systematic overhaul of the procedures for the collection of race and ethnicity data as (1) assessment and evaluation, (2) infrastructure modification, (3) training and education, (4) implementation and response to results, and (5) acknowledging limitations and lessons learned."

In Phase I of the study, researchers conducted an assessment of the current data collection process, interviewed staff and assessed the system they were studying and "determined that the system had no reliable or cohesive plan for the collection of REL (race, ethnicity and language preference) data." Additionally, their institution's Office of Diversity and Inclusion performed client and staff interviews and a group discussion on the targeted site. They used this data to develop training aimed at addressing issues found and kept the conversations moving forward to monitor progress in the latter stages of the study.

Phase II of the study required modifications to the data system infrastructure. While there was a core system, the entire system was reliant on the input from seven separate registration systems. These systems were the result of mergers and acquisitions of legacy systems. The data system teams were actively involved in developing "new race and ethnicity data entry fields, mak[ing] entries mandatory, and add[ing] more descriptive identities" for their clients, since some of the clients visited multiple

locations. The collection went beyond the race and ethnicity categories outlined in OMB 1997 requirements, before SDF-15's new rules. The researchers emphasized the need to be consistent and how that would positively affect the scalability of their project.

Phase III of the study was centered on training and educating staff on the best data collection processes related to health equity based on the findings from Phase I. Unconscious bias education was also implemented for front-line staff to emphasize the importance of reliable data in achieving health equity and social justice. The learning objectives for Phase III included:

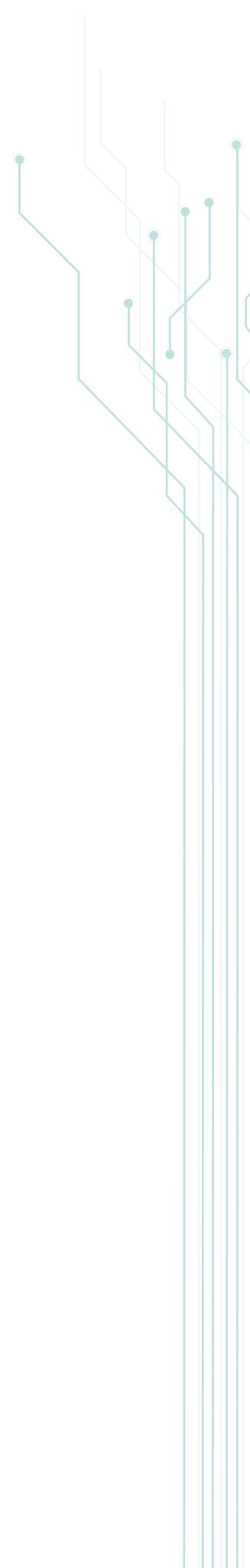
- Defining race and ethnicity.
- Establishing the purpose of collecting the data and how to properly use the race/ethnicity data collection form.
- A review of the location of the data fields in each registration system.
- The creation of a frequently asked questions document to be accessed by both patients and staff. Training was conducted using on-demand training and development platforms, instructor-led webinars and in-person classroom instruction. These efforts resulted in the successful implementation of the race and ethnicity data collection process.

Phase IV focused on evaluating the feedback of the effectiveness of the training and data collecting procedures. By reviewing the feedback, staff were able to identify any deficiencies in need of improvement. The study found that empowering patients and emphasizing their role in helping provide equitable medical care yielded positive results among patients and staff. These efforts resulted in a sustained increase in race and ethnicity data fields and a decrease in the number of unidentified patients over five years. By year five, 90% of patient race and ethnicity data was captured and the number of unidentified patients dropped by 76%.

Phase V of the study focused on the limitations and lessons learned. A successful relationship with the IT department proved to be a valuable component in improving the patient data collection improvement process and additional elements, such as leadership and organizational accountability. The study revealed that one of the areas of improvement when collecting patient data was identifying registration and scheduling leaders. The IT department identified a lack of consistent lead personnel for each registration system. This persuaded the IT department to create the registration governance group that meets consistently to coordinate relevant work updates across multiple health systems. Additionally, the study noted that the collection of more granular demographic data, including the addition of patient identity categories based on the demographics of the community, is essential for further data improvement. Examples of additional demographic categories can include gender identity, language preference, housing status and other intersectional variables. However, new variables require additional data fields that would require implementation at a later time and not for the scope of this project.

In conclusion, the research established that accurate race and ethnicity data essential in creating effective, equitable care is possible. The patient registration data collection improvement process challenged the standard data collection recommendations from traditional literature. The patient registration data collection improvement process incorporated improving system workflows, communication strategies and procedural training.

This study implemented system-wide interventions that improved the process of capturing race and ethnicity data to achieve a greater than 90% patient self-identification capture rate. Improving the collection of accurate race and ethnicity data led to the development of core equity dashboards across health systems that covered race, ethnicity, age, gender and other demographic variables.





Recognizing patient hesitancy toward providing demographic data, such as race and ethnicity, is also essential. Although concerns may arise, it is necessary to establish rapport and emphasize the patient's role in creating equitable care, which proved beneficial. The study's success proved race and ethnicity data to be critical in eliminating health injustices and creating equal access to health care.

SUCCESS STORIES

Many states have been successful in utilizing available demographic data to better deliver services, inform policymaking decisions and better understand complex issues. The Pew Charitable Trust's report, "[How States Use Data to Inform Decisions](#)," offers success stories from states that leveraged demographic data to achieve results.

SERVICE DELIVERY

Officials in Virginia were facing a problem as the number of homeless veterans was increasing. To better understand the scope of the issue and how to better deliver services to veterans, Virginia officials [leveraged existing data on veteran homelessness to deliver services more efficiently](#). The Federal Department of Veterans Affairs, as well as regional and community organizations, began sharing data to deliver services, evaluate strategies and monitor outcomes. Through this, they were able to identify gaps in service delivery and better target services to fill those gaps. As a result, [Virginia functionally ended veteran homelessness](#) by linking and analyzing information from 16 disparate data sets and housed 2,737 veterans from October 2014 to December 2016.

UNDERSTANDING COMPLEX PROBLEMS

Indiana was facing one of the [highest infant mortality rates](#) in the United States. State officials realized that to solve the problem, they needed to better understand what was going on. [Analysts combined data](#) from agencies and public sources to see the full scope of the available data. By analyzing the data, Indiana officials found that young mothers on Medicaid were not receiving the recommended number of prenatal visits. With the new information gained from the study, the Indiana Department of Health initiated [a statewide effort to specifically address this issue](#). Leveraging the available data allowed state officials to understand what was going on in greater detail and deploy services that could target this complex problem.

LEVERAGING U.S. CENSUS BUREAU'S AMERICAN COMMUNITY SURVEY (ACS) DATA

The YMCA wanted to better understand characteristics of the neighborhoods they serve, so they developed the [Community Insights](#) tool, which utilizes both YMCA's internal data and available ACS data. Demographic data available in the ACS data helped the YMCA better understand the unique needs of the local communities they serve. When the West Cook YMCA Oak Park, Illinois, received a \$1.27 million grant to tackle chronic illness among community members, Community Insights was used to help West Cook YMCA identify the best locations for the program by utilizing ACS demographic data including total population, poverty rates and Medicaid coverage. This data allows the West Cook YMCA to better target their grant spending and reach more than 20,000 of the 52,000 residents who live in the area who have or who are at risk of diabetes, obesity and high blood pressure.

Insights from ACS data allowed the [American Psychological Association \(APA\)](#) to secure needed funding for programs that are working to strengthen the mental health workforce and meet the needs of communities. The APA used ACS data to analyze the psychology workforce and education pipeline. The ACS data allowed the APA to understand the demographic composition of the psychology workforce, how it has changed over the years and whether it is adequate to address population health needs. APA used the information to advocate for increased funding for psychology workforce programs and helped secure a \$3 million increase for psychology education and training to build a more robust and diverse workforce.

8. Advancements in Demographic Data Collection

In recent years, rapid advancements in technology have revolutionized how demographic data is collected, analyzed and utilized. This section explores the trends shaping demographic data collection, focusing on the integration of emerging technologies, artificial intelligence, machine learning and predictive analytics.

EMERGING TECHNOLOGIES IN DEMOGRAPHIC DATA COLLECTION

Emerging technologies such as mobile devices, internet of things sensors and social media platforms have transformed the landscape of demographic data collection. [Mobile applications](#) enable real-time data collection, allowing researchers to gather information efficiently and accurately. [Internet of things sensors](#) provide insights into urban dynamics, environmental factors and population movements. Social media platforms offer valuable demographic insights through user-generated content and interaction patterns. These technologies offer unprecedented opportunities to capture demographic data with enhanced granularity and timeliness.



INTEGRATION OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence and machine learning algorithms play a crucial role in processing and analyzing vast amounts of demographic data. [Natural language processing](#) techniques extract valuable insights from unstructured data sources such as social media posts, online forums and news articles. Image recognition algorithms enable demographic inference from visual data, facilitating the analysis of population dynamics in urban areas. From that data, machine learning models can identify patterns and correlations within demographic datasets, uncovering hidden relationships and trends that traditional statistical methods might overlook. The integration of AI and machine learning enhances the efficiency and scalability of demographic data analysis, enabling organizations to derive actionable insights more effectively.

PREDICTIVE ANALYTICS AND FORECASTING

[Predictive analytics](#) leverages historical demographic data, AI algorithms and statistical modeling techniques to forecast future trends and scenarios. By analyzing past demographic patterns and socio-economic indicators, predictive analytics can anticipate changes in population dynamics, migration patterns and consumer behavior. These insights enable governments, businesses and organizations to formulate proactive strategies and policies, mitigating potential risks and capitalizing on emerging opportunities.

According to [SAS](#), a leader in AI and data management analytics, predictive analytics can:

- Detect fraud by improving pattern detection to identify criminal behavior.
- Optimize marketing campaigns by determining responses and promote opportunities.
- Improve operations by forecasting inventory and capacity and managing resources.
- Reduce risk by understanding client behavior and developing models to predict risk-related issues, like the need for financial counseling.
- Moreover, predictive analytics allow stakeholders to allocate resources efficiently, address demographic challenges and optimize service delivery based on anticipated demand.

CHALLENGES AND CONSIDERATIONS

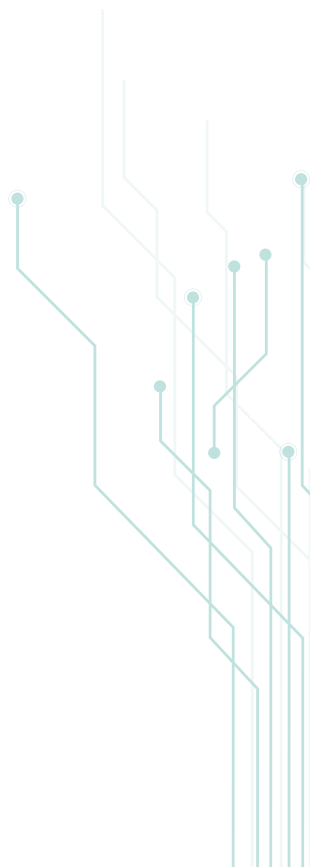
While the integration of emerging technologies, AI and predictive analytics offers significant benefits for demographic data collection and analysis, several challenges and considerations must be addressed. Privacy concerns surrounding the collection and utilization of personal data require robust data protection measures and ethical frameworks. As Criado (et al.) discussed, there are added issues with digital discrimination in the formation and assumptions used to build algorithms. [Bias and discrimination in AI algorithms](#) can perpetuate existing inequalities and distort demographic insights, necessitating algorithmic transparency and fairness. It is important to note that bias does not necessarily mean the social group is being treated poorly or put at a disadvantage; it simply means that they are being treated differently from the standard. Additionally, ensuring data accuracy, reliability and representativeness remains paramount to derive meaningful insights and make informed decisions.

The evolution of demographic data collection is driven by the convergence of emerging technologies, AI integration and predictive analytics. These trends enable stakeholders to capture, analyze and leverage demographic insights with unprecedented depth and precision. By embracing innovation and addressing associated challenges, organizations can harness the power of demographic data to inform policymaking, drive economic development and enhance societal well-being in an increasingly complex and dynamic world.

9. Conclusion

This paper provides an overview of the challenges, benefits and ethical considerations in demographic data collection. It is designed to be a tool teams can reference as they develop their data collection process. When demographic data is collected and leveraged correctly, it can serve as a powerful tool for decision-making across a range of sectors. It can be especially useful for local, state and federal agencies as a driver of evidence-based policymaking. By accurately capturing information on population characteristics such as age, race, income and education, governments and non-governmental agencies can better understand their client's needs, allocate resources equitably and formulate policies that address diverse societal challenges. Demographic data not only ensures fair representation and access to services but also fosters transparency and accountability in decision-making processes.

Preparation of this item was fully funded by the United States Department of Labor's Office of Apprenticeship. This item does not necessarily reflect the views or policies of the U.S. Department of Labor, nor does mention of any trade names, commercial products or organizations imply endorsement by the U.S. Government.



ADAPTAC



10. Resources

- BrckaLorenz, A., Haeger, H., & Zilvinskis, J. (2014, May). Categorizing identities: Race, gender, disability, and sexual orientation. Association for Institutional Research Annual Forum.
- Brown, K., & Johnson, M. (2021). Social Media Analysis for Demographic Research: Opportunities and Pitfalls. *Journal of Social Research*, 38(4), 345-367.
- Catalogue of Workforce Information Sources: Decision Making Assistance for Regional Economic Development. <https://files.eric.ed.gov/fulltext/ED508184.pdf>
- Chen, X., & Zhang, L. (2019). Machine Learning for Demographic Data Integration: Challenges and Opportunities. *Journal of Data Science*, 24(2), 189-211.
- Cote, C. What is Predictive Analytics: 5 Examples, Harvard Business School. October 2021. <https://online.hbs.edu/blog/post/predictive-analytics>
- Criado N. and Such J. M., "Digital discrimination" in *Algorithmic Regulation*, Oxford, U.K.: OUP, 2019.
- DeMeester, R., Mahadevan, R., Cook, S., Martin, J., Tuck, K., Patel, S., & Durkin, E. (n.d.). Using Data to Reduce Disparities and Improve Quality. *Advancing Health Equity*. https://www.chcs.org/media/Using-Data-to-Reduce-Disparities-2021_Final.pdf
- Faster Capital. Case Studies of Businesses that used demographic data effectively. <https://fastercapital.com/keyword/studies-success-stories-studies-studies.html>
- Ferrer, T., Nuenen V., Such J.M., Coté M. and Criado N., "Bias and Discrimination in AI: A Cross-Disciplinary Perspective," in *IEEE Technology and Society Magazine*, vol. 40, no. 2, pp. 72-80, June 2021, doi: 10.1109/MTS.2021.3056293.
- Fischer F, Kleen S. Possibilities, Problems, and Perspectives of Data Collection by Mobile Apps in Longitudinal Epidemiological Studies: Scoping Review. *J Med Internet Res*. 2021 Jan 22;23(1):e17691. doi: 10.2196/17691. PMID: 33480850; PMCID: PMC7864774.
- Husband, Gary. 2020. "Ethical Data Collection and Recognizing the Impact of Semi-Structured Interviews on Research Respondents" *Education Sciences* 10, no. 8: 206. <https://doi.org/10.3390/educsci10080206>
- <https://www.federalregister.gov/documents/2024/03/29/2024-06469/revisions-to-ombs-statistical-policy-directive-no-15-standards-for-maintaining-collecting-and>
- <https://www.bls.gov/opub/mlr/2014/article/pdf/the-development-of-questions-on-disability-for-the-current-population-survey.pdf>
- Hughes J.L., (et al.). Rethinking and Updating Demographic Questions: Guidance to Improve Descriptions of Research Samples *Journal of Psychological Research* Fall 2016 Vol 21 Issue 3
- Jones, A., & Patel, R. (2020). Leveraging Administrative Data for Population Research: Opportunities and Challenges. *Population Studies*, 44(3), 267-289.
- Jones, N., Marks, R., Ramirez, R., & Rios-Vargas, M. (n.d.). 2020 Census Illuminates Racial and Ethnic Composition of the Country. *Census.Gov*. Retrieved June 21, 2024, from <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>
- Kirst M, (et al.). Sociodemographic Data Collection for Health Equity Measurement: a Mixed Methods Study Examining Public Opinions. *International Journal for Equity in Health* 12, Article number: 75 (2013) <https://equityhealth.biomedcentral.com/articles/10.1186/1475-9276-12-75>

Kumar V, (2021) Defining Cultural Data Science. <https://discovery.ucl.ac.uk/id/eprint/10141034/>

Lin, L., & Ginsberg, A. (2023, December 15). American Psychological Association Uses ACS Data to Identify Need for Mental Health Services, Education, and Training. Census.Gov. <https://www.census.gov/programs-surveys/acs/about/acs-data-stories/psychologists.html>

Madden, M., & Rainie, L. (2015, May 20). Americans' Views About Data Collection and Security. Pew Research Center. <https://www.pewresearch.org/internet/2015/05/20/americans-views-about-data-collection-and-security/>

O'Neill, O. Some limits of informed consent. *J. Med. Ethics* 2003, 29, 4–7.

Pew Charitable Trusts. (2018) How States Use Data to Inform Decisions.

Pillai, D., & Published, S. A. (2024, April 30). Revisions to Federal Standards for Collecting and Reporting Data on Race and Ethnicity: What are They and Why do They Matter? KFF. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/revisions-to-federal-standards-for-collecting-and-reporting-data-on-race-and-ethnicity-what-are-they-and-why-do-they-matter/>

Research to Improve Data on Race and Ethnicity. Census.Gov. Retrieved June 21, 2024, from <https://www.census.gov/about/our-research/race-ethnicity.html>

Serrano, M.-A. (2023, December 15). The YMCA's Community Insights Tool Uses ACS Data to Maximize Impacts for Communities. Census.Gov. <https://www.census.gov/programs-surveys/acs/about/acs-data-stories/ymca-community-insights-tool.html>

Smith, J. (2019). Survey Methods for Demographic Data Collection: Challenges and Best Practices. *Journal of Population Research*, 36(2), 123-145.

Teclaw, Robert & Price, Mark & Osatuke, Katerine. (2012). Demographic Question Placement: Effect on Item Response Rates and Means of a Veterans Health Administration Survey. *Journal of Business and Psychology*. 27. 10.1007/s10869-011-9249-y.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5), 859–883. <https://doi.org/10.1037/0033-2909.133.5.859>

Vega Perez R D, Hayden L, Mesa J, et al. (January 05, 2022) Improving Patient Race and Ethnicity Data Capture to Address Health Disparities: A Case Study From a Large Urban Health System. *Cureus* 14(1): e20973. DOI 10.7759/cureus.20973

von Schrader, S., Malzer, V., Erickson, W., & Bruyère, S. (n.d.). Emerging Employment Issues for People with Disabilities. Cornell University, ILR School, Employment and Disability Institute.

Wang, Y., & Li, H. (2018). Geospatial Data Analysis for Demographic Research: Methods and Applications. *Demography*, 55(1), 123-145

Wen, K. 5 Key Data Compliance Regulations to Know for 2022. SimpleLegal, May 2022. <https://www.simplelegal.com/blog/data-compliance-regulations>

