Developing a 21st Century Census Curated Data Enterprise:
A Bold New Scientific Approach for Official Statistics

Stephanie S. Shipp,
https://orcid.org/0000-0002-2142-2136
sshipp919@gmail.com
sshipp@iastate.edu

Census Curated Data Enterprise Team:
Zhengyuan Zhu, Steve Jost, Joan Naymark, Ken Prewitt, Joseph Salvo,
John Thompson, Matthew Snipp, Erika Becker-Medina

March 7, 2025

.

UNIVERSITY *of* VIRGINIA

BIOCOMPLEXITY INSTITUTE

We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win, and the others, too.
President John F. Kennedy, speech at Rice University, September 12, 1962

Abstract

Questions we now seek to answer often go beyond what traditional censuses or surveys alone can hope to address. A prominent example concerns disparities in the availability and quality of data for population groups and communities. Limited sample sizes and large margins of error mean many official statistics lack demographic details entirely or fail to represent small population groups and sparsely populated rural communities accurately. Privacy concerns further limit access to disaggregated data, preventing communities from fully understanding their challenges and solutions. This lack of data, or inaccurate data when it exists, makes identifying, measuring, and reducing disparities challenging. Similar issues arise across disciplines and sectors, prompting researchers and policymakers to seek novel data sources -such as unstructured text scraped from the internet and linkages among state and federal administrative records- to supplement structured numeric data from traditional sources, to fill data gaps and improve the quality and usefulness of statistical products. However, these efforts are often ad hoc, siloed, and duplicative, resulting in sub-optimal solutions.

That is the impetus for developing the 21st Century Curated Data Enterprise (CDE), an innovation in data science designed to create statistical products from diverse data types and build the infrastructure to support them. The CDE, as the name implies, features an end-to-end curation model that captures the entire statistical product development process. Motivated by the US Census Bureau's needs, this initiative supports Census Bureau scientists and data users to shift from a focus on single survey-driven data elements to a framework that integrates and curates multiple data sources. This integration aims to produce better social and economic measures, enhancing social science measurement, research, planning, and many other important applications.

This report describes and summarizes our research to address two tasks requested by the Census Bureau:

Task 1: Identifying and Characterizing Stakeholder Communities and

Task 2: Developing a research agenda.

Just like putting humans on the moon, the Census Bureau is setting clear objectives, dedicating resources to meet the objectives, and pursuing internal and external collaborative efforts to socialize and produce statistical products in a bold new way (Santos 2022, 2023, 2024; Keller 2024). This new approach aims to produce better measures of people, places, and economies for a wide range of important purposes and uses.

Table of Contents

Developing a 21st Century Census Curated Data Enterprise:
A Bold New Scientific Approach for Official Statistics

A. Defining the Curated Data Enterprise (CDE)

Two centuries ago, when the framers of the US Constitution laid the cornerstone for the federal statistical system, they could not have imagined the complexity of questions future generations would ask or the variety of data sources available to address them. When the Constitution was written, with its requirement for a decennial census, counting the population and apportioning state seats in the House of Representatives were among the most urgent tasks before the young nation. Now, 237 years later, the decennial census continues to serve its original purpose – but the purposes and uses of US Census Bureau data need to support have expanded. With the tremendous expansion of data availability, there is an opportunity to combine all types of information to create statistical products that more fully address user needs. A two-century traditional view of Census Bureau official statistics falls short of informing the future state of our people, places, and the economy. Our initial work was described in *Developing a 21st Century Census Curated Data Enterprise: A Bold New Approach to Developing Federal Statistics* (Keller et al. 2022). This report describes the accomplishments since this report was published in Spring 2022.

Indeed, the questions the country now demands answers to go beyond what the decennial census and its offshoot, the American Community Survey, alone can hope to address. Even with the multitude of other surveys conducted by today's Census Bureau, researchers and policymakers find themselves looking to additional sources of data – from linkages of federal and state administrative records to unstructured text documents scraped from the internet – to supplement traditionally designed data to address such topics as the extent to which nursing homes and communities are prepared for extreme climate events. Combining data sources can fill data gaps and improve the quality and relevance of the statistics produced. Examples include creating evidence-based economic affordability standards, new policy-relevant indicators (for example, estimates of migration consequent to wildfires and other natural disasters), and state-based integrated data systems on young children.

We are collaborating with the Census Bureau to propose a bold new approach to develop statistical products for the future by focusing on their purpose and uses. This approach begins with the end in mind by engaging with stakeholders to identify their information needs to reach their objectives and, from there, shape the statistical products to be developed. The scaffold for this approach is a **21st Century Census Curated Data Enterprise (CDE)** that aligns with the Census Bureau's modernization and transformation of its enterprise systems (see Exhibit 1). By linking the ecosystem across activities, data, and frames, the Census Bureau is creating the enabling technologies that make this approach possible.

Exhibit 1. Curated Data Enterprise Framework

This issue, or aspiration, is not unique to the Census Bureau and is being embraced across the globe by official statistics agencies. These challenges can be addressed through the CDE, a scientific framework and technology designed to enable the creation of statistical products from all data types by constructing the infrastructure to support their rapid development. Within the CDE, the development of statistical products will start with understanding stakeholder questions and focus on building data products to support stakeholder purposes and uses by using and repurposing all data types – designed surveys and censuses, public and private administrative data, opportunity data scraped from the internet, and procedural data (Keller et al. 2020). Once existing data

have been exhausted and gaps identified, then consideration can be given to new data acquisition and new model development.

The choice of the word "curated" in the CDE is deliberate. For the CDE to succeed in supporting the agile and rapid development of useful statistical products, careful end-to-end curation, thoughtful stakeholder interactions, data choices, and product development and dissemination are critical. Curation ensures the reuse of data, research, and stakeholder insights and concerns to support the creation of the next set of statistical products.

In Fall 2022, a Census Bureau working group crafted a draft definition for data curation that includes the above ideas. Data curation encompasses efforts to support preserving and adding value to data, including:
- organizing data to facilitate discovery and provide access.
- documenting data to enable the reuse of the data in scientific and programmatic research.
- enhancing the value of the data ecosystem a) through linkages between datasets and b) by mapping the network of interconnections between datasets, research outputs, researchers, and institutions.

The CDE aligns with the Census Bureau's modernization and transformation (Thieme 2022) while maintaining the fundamental responsibilities of statistical agencies (OMB 2023, NASEM 2023, NASEM 2024). The development of the CDE supports the Census Bureau's *Statistical Products First* initiative by identifying stakeholders' purposes and uses and the capabilities needed for the Census Bureau to build the CDE.

The CDE provides a framework to address many of the challenges that statistical agencies face today, leading to new, innovative methods of integrating a variety of data sources into official statistical products tailored for specific purposes and uses. The CDE development is coming together rapidly due to collaboration with Census Bureau leadership. The concept of the CDE is important. A curation pipeline will likely be built in the Census Bureau's IT Enterprise System. ***The CDE is both a framework to guide the creation of innovative data products for use cases and a scientific technology that can be built as an independent end-to-end curation system or embedded in a larger system.***

An analogy to developing the CDE to produce better data for the nation is NASA's commitment to putting humans on the moon. The success of the Apollo program not only accomplished the immediate goal of landing humans on the Moon but also spurred technological and scientific advancements with lasting impacts on society. It is a testament to what can be achieved through clear objectives, resources, and a collaborative effort. The same applies to the Census Bureau's initiative to create new statistical products that use new data science methods and multiple data types.

The Alfred P. Sloan Foundation provided complementary support to achieve the objectives.

B. Tasks and Deliverables

This section aligns with the reports requested in the US Census Bureau memorandum.

Task 1: Identifying, Characterizing, and Engaging with Stakeholder Communities.

Task 2: Developing a Research Agenda.

In Task 1, we highlight the themes heard across Listening Sessions on Purposes and Uses, data wish lists, and ideas for tools and infographics. Our interactions with Census Bureau experts, researchers, and other experts are discussed in Appendices 1, 4, 5, 6, and 7. The themes for each Listening Session individually are described in Appendix 2.

In Task 2, we summarize the demonstration and art of the possible Use Cases and our review of State Data Centers. More details are provided in Appendix 3.

Task 1: Identifying, Characterizing, and Engaging Stakeholder Communities through Listening Sessions.

Task 1 activities include creating and refining the stakeholder taxonomies, conducting Listening Sessions, and holding discussions with experts to socialize the **Curated Data Enterprise (CDE)** and Statistical Product First approach, to seek stakeholder ideas about Purposes and Uses and other insights relative to the creation of new statistical products. We also discuss a method for prioritizing purposes and uses.

In the first phase of this research, we defined six stakeholder groups:
- Congress and the Legislative Branch.
- Public Policy Community.
- Business and Commerce Community.
- Nongovernmental organizational (NGO) Community.
- Researcher Community.
- Media.
- General Public.

Through our Listening Sessions and discussions with experts, we expanded the stakeholder taxonomy to describe stakeholder roles related to data needs on resilience in more detail.
- Nongovernmental organizational (NGO) Community: We talked to community groups involved in children's health, community health, and resiliency of downtowns and other areas susceptible to natural disasters.
- Researcher community – University centers, research groups, and federal agencies that conduct research provide data or support research on resiliency, which is broadly defined.

These groups included:

- Advocacy: children, health, community solutions, downtowns, housing recovery.
- Federal agencies: NOAA, NIST, FEMA.
- Federal Reserve Banks.
- Researchers at Federal Statistical Research Data Centers (FSRDCs).
- State and local data users.
- University research centers.

Themes and Findings from Listening Sessions and Discussions with Experts.
As noted above, the CDE framework starts with understanding stakeholder questions and focuses on creating data insights to support stakeholder purposes and uses. The overarching theme for our listening sessions was seeking stakeholder purposes and uses to identify statistical products that inform their needs. We also asked stakeholders about how to achieve data fairness, noting that many of our official statistics do not provide information by demographic characteristics, or if they do, they do not accurately count underserved populations.

Data for many cultural communities lack sufficient granularity to identify subgroup variation within a larger group. Small population groups and those in small geographic areas lack accuracy due to small sample sizes and large margins of error. Disaggregated data are not readily available to communities to understand their challenges and solutions. This lack of data, or inaccurate data when it exists, makes identifying, measuring, and reducing disparities challenging. Robert Santos, the US Census Bureau Former Director (2021-2025), notes in his blog that inclusion also applies to listening to community voices (Santos 2022). Thus, we engaged in discussions with stakeholders and experts about identifying statistical products that answer their questions.

Exhibit 2 summarizes the numbers of participants, observers, and Census CDE working group and Census Bureau employees at each session.

Exhibit 2. Listening Sessions and Conference presentations by data and counts.

Listening Sessions by organization	Date	Number of participants	Number of Observers	Number Census WG
Use Case Requirements	6/9/2022	6	3	3
Curation across Disciplines	6/22/2022	10	4	4
Local Data Partnerships	12/6/22	11	5	10
Third-Party Data (economists from large companies and the Federal Reserve Banks)	1/17/23	10	8	10
Undercount of Young Children	9/15/23	6	2	10
National Healthy Start Association	12/15/23	7	3	11
Community Health Advocates	12/21/23	5	3	4
Community Solutions	1/16/24	11	11	9
National Low Income Housing Consortium (NLIHC) Disaster Housing Recovery Research Consortium Coalition (DHRRC)	5/15/24	23	4	6
International Downtown Association	6/12/24	15	3	2
Colorado State University NIST Center of Excellence	6/18/24	20	0	2
National Low Income Housing Consortium (NLIHC), Disaster Housing Recovery Research Consortium Coalition (DHRRC)	10/22/24	30	0	2

Listening Sessions held at conferences	Date	Number of participants	Number of Observers	Number Census WG
American Community Survey (ACS) Users Conference	5/15/23	80	-	4
Association of Public Data Users (APDU)	7/26/23	60	-	2
Total		**268**	**46**	**79**

We also presented our research at the Federal Committee on Statistical Methodology (FCSM) conferences in 2022 and 2023 and the Applied Demography Conference in 2024, introducing the CDE to many more people.

The main report summarizes ideas for purposes and uses, tools, infographics, and data sources heard *across* the Listening Sessions. In Appendix 2, we provide summaries for each Listening Session and discussion with experts.

In preparation for these Listening Sessions and to expand our knowledge, we talked to experts and read the literature. In the first set of Listening Sessions held in the Fall of 2021, we identified stakeholders individually (Keller et al., 2022). We continued this approach for the first two Listening Sessions under this Memorandum of Understanding (MOU). We switched gears for the remaining Listening Sessions by identifying relevant organizations. These organizations would invite their leadership and members to the Listening Session. This approach extended our outreach to new stakeholders and individuals.[1]

Findings across Listening Sessions with external stakeholders

After introductions, the Listening Sessions started with an overview of the Curated Data Enterprise (CDE) and the Statistical Products First approach. We then had two rounds of questions, slightly tailored to each group. Members of the CDE Team (see cover page) and Census Bureau employees participated in these different roles leading discussions or observing, thus providing a variety of voices and styles. The session ended with a round robin to gather last thoughts and insights. We asked the questions presented in Exhibit 3.

Exhibit 3. Listening Session Discussion Questions.

Discussion 1: Purposes and Uses.
  ➢ What are major purposes and uses the statistical products need to support? (What statistical products would help you?)
  ➢ What would add value to your statistical products?
  ➢ What modalities of statistical product dissemination support data user needs? (What tools would help you find and use data?)

Discussion 2: Fair implementation.
  ➢ How do we elicit information about purposes and uses fairly?
  ➢ Who should participate in the development of products?
  ➢ How do we ensure that the development of statistical products happen fairly?

---

[1] Robin Bachman, Bethany DeSalvo, and others at the US Census Bureau provided ideas and made introductions to organizations.

The findings from the Listening Sessions are grouped by topic:
- Purposes and uses.
- Data Wish List.
- Custom Geographies Data Wish List.
- Health Data Wish List.
- Small Area Estimation Data Wish List.
- Tools.
- Infographics.
- Eliciting information from stakeholders.

Purposes and Uses

> As part of our transformation and modernization, we're reimagining the way we produce statistical data. We've commenced a "Statistical Product First" approach that specifically focuses on expressed public needs. Under this new approach, we engage with the public to identify the data *purposes and uses* that people need. This drives the development of the statistical products. And this outreach is iterative to ensure continuous improvement and clear understandings.
> *Robert L. Santos, US Census Bureau Director's [Blog](#), February 26, 2024*

At the Listening Sessions, we introduced the Curated Data Enterprise (CDE), noting that it offers unprecedented opportunities for capturing the ever-changing economic and social landscape. From managing pandemics and natural disasters to adapting to shifts in the workforce, real-time and geographically granular data is essential for addressing critical questions raised by policymakers, media, analysts, researchers, planners, advocates, and the public. The CDE is purposefully dedicated to leveraging Census Bureau data, anticipating and overcoming the challenges of uncovering, integrating, and creatively using and disseminating data and statistical products in novel ways.

Through Listening Sessions, we identified purposes and uses for which stakeholders need data to address their questions comprehensively. In this phase of our work, we primarily focused on advocacy and resiliency. Those themes, along with others, surfaced through these sessions. We grouped the Purposes and Uses by topics framed as questions.
- Demographic and Business.
- Community Characteristics and Social Determinants of Health.
- Young Children.
- Downtowns.

Demographics and Business Purposes and Uses

Hybrid/Remote Work Patterns: What are hybrid and remote work patterns? This would inform workplace policies and economic planning. (Third Party-Data, International Downtown Association)

Personal Income Measurements: Can the accuracy of measuring personal income across various demographics and geographies be improved? The focus is to improve policy targeting and economic analysis. (Third Party-Data.)

Supply Chain Analysis: What are supply chain dynamics at the local level? A consistent theme was the need for insights into supply chain dynamics at a local level. This highlights a broader interest in understanding how goods and services flow through economies and the impact of disruptions. (Local Data Partnerships, Third Party Data.)

Access to Finance and Capital: How do small businesses access capital? The pandemic underscored the importance of understanding how businesses, particularly small and local ones, navigate financial challenges and access capital. This is crucial for economic resilience and recovery planning. (Local Data Partnerships, Third Party Data, others.)

<u>**Community Characteristics and Social Determinants of Health Purposes and Uses**</u>

What is the relationship between the social determinants of health and race on health outcomes at the community level? Participants found earlier Centers for Disease Control and Prevention (CDC) data that highlighted how race impacts health outcomes to be tremendously useful for connecting the dots between social determinants and health outcomes. (Early Childhood Integrated Data Systems, Healthy Start.)

What are the characteristics of high-need communities? What challenges do they face accessing services? High-need communities require support for food, transportation, healthcare, and other services. Can the integration of services help these groups to thrive? (Community Health Advocates.)

Where are service deserts? What distances and challenges do children and older people face in obtaining health and related services? Are there geospatial data products at the block level to help identify service deserts and effectively plan resource allocation? (Community Health Advocates.)

What are the social determinants of health for the uninsured population? What are their needs and challenges? (Community Health Advocates.)

What are older Black people's demographic characteristics and health status in a specific census tract and block? What are the community assets and gaps in delivering health services to a community? Can the ACS, CDC data, and other data sources be integrated to address these questions? (Community Health Advocates)

Who is disabled according to the federal definition of disability? (One expert said that the ACS captures about half of the disabled population.) (Community Solutions – *see Appendix 1 for details).* People with disabilities often don't identify as having a disability. They say that they have a chronic health condition. Advocates who advance policy and program changes need accurate data.

Who has access to transit versus who uses public transit? What is the relationship between driver's license suspensions and the use of public transit? What are the reasons for suspensions, e.g., bad driving, unpaid debts, etc.) (Community Solutions)

Do food banks reduce the number of food-insecure households? What are other sources of food that food-insecure households get? (Community Solutions)

What are local literacy rates, school enrollment, and employment trends? These data are needed for planning and development purposes to address educational needs and to understand employment trends for better workforce planning. **(Community Solutions)**

What types of legal problems do low-income (poor) people have? (Community Solutions.)

- Note: The Legal Services Corporation (LSC) is federally funded through congressional appropriations. LSC periodically runs the **Justice Gap Survey** to measure the prevalence of various legal problems, especially for the poor. See Appendix 1 for more details.

How does redlining, the history of racial discrimination in appraisals, mortgages, etc., affect communities' resiliency today? (NLIHC DHRRC)

Does Green infrastructure save money? What are the timelines? (NLIHC DHRRC)

Young Children Purposes and Uses

Undercount of Young Children (Community Solutions, Health Advocates.)

- "Half of that problem is that people think that the census form is only for adults. They just leave children off. They are not aware of who should be included."
- One way to fill the gap is to use Administration for Children and Families (ACF) administrative data. This involves negotiating for the data with each state.
- Data from daycare centers, pediatric clinics, and community health programs could provide alternative counts or insights into the population of young children.
- State Early Childhood Integrated Data Systems (ECIDS) and State (K-12) Longitudinal Data Systems (SLDS) are another source of data.

Can State Early Childhood Integrated Data Systems improve the accuracy and timeliness of counting and characterizing young children? How do public health estimates of children in poor health change with more accurate data? (Early Childhood Integrated Data Systems, Healthy Start.)

Can the quality of the count of young children be enhanced using SNAP and WIC data? How does the acquisition and distribution of funds for childcare change with a more accurate count of young children? "What keeps us up at night is worrying about how accurate the denominator is – the count of young children." (Early Childhood Integrated Data Systems.)

What is the effectiveness of early intervention programs, e.g., Are children's developmental milestones on target? Can Early Childhood Integrated Data Systems provide comprehensive insights into the health, education, and social service interactions of young children? (Early Childhood Integrated Data Systems, National Healthy Start Association.)

Pre-Workforce Status: What is the status of individuals before they enter the workforce, e.g., school, military, etc.? This information can inform policies around education, training, and veteran reintegration. (Local Data Partnerships.)

Downtowns Purposes and Uses

How many people work in downtown areas? Is it possible to differentiate visitors from workers? Potential data sources: placer.ai, arrivalist.com.

What is the distribution of current and emerging talent? What is the path from attainable to the aspirational job? Are certifications, community college degrees, and other training credentials indicators of current and emerging talent supply and demand? Potential data are NLx.

What are the characteristics of Activity Districts in a downtown area, e.g., what do planners need to know if deciding to build a stadium? How does this change the economic and social characteristics of an area? Does it attract businesses, new residences, or other spillovers (positive and negative)? What is the impact radius (1/4 mile, ½ mile, or more)?

What are the characteristics of resilient downtown areas? Some ideas include the presence of multi-use buildings (commercial, retail, hotels, residences); schools, daycare, parks, and sidewalks to attract families; walkability, reduced car usage, increased public transportation use; bike lanes; community engagement – use of retail, attendance at sports and concerts, etc.; and inclusive urban areas. The *State of Downtown Reports* are available from the International Downtown Association.

Who is moving into downtown areas? Who is moving out? Where are they coming from/going to? Are families moving to downtown areas? What are household compositions and ages? Potential data sources: placer.ai, cell phone data, and Arrivalist (https://www.arrivalist.com/).

How many people are unhoused? What are trends (changes) over time? What types of housing are available for the unhoused? What services are available for the unhoused?

What kinds of retail should a downtown area develop? What are population characteristics within a 5-mile radius of downtown, e.g., percent with a car, have three or more people in their household, have high incomes? What makes the area attractive to develop?

Do people stay in their neighborhood or go outside their neighborhood to shop (called sales leakage)? Compare spending by category, such as restaurant revenue, to actual gross revenues. Are dollars leaking to other areas?

What are clusters of downtowns that are similar? For example, what are five other downtown areas that are most similar to St. Louis? Create an index to facilitate comparisons. For example, the two downtowns are not similar if St. Louis is 100 and Miami is 45. Analysis could initially be done at the Metropolitan Statistical Area (MSA) level and then downtown areas. IDA's Value of Downtowns[2] is a useful tool to build on.

Data Wish Lists

Many stakeholders use Census Bureau and other data to create their products. They discussed their wish list for new data sources, changes to existing data sources, tools, and infographics to support their work.

These ideas are grouped into the following data wish lists:
- American Community Survey and related data.
- Custom geographies.
- Health and health insurance data.
- Property insurance data.
- Small area estimates and synthetic data.

American Community Survey (ACS) and Related Data: Ideas from Listening Session Stakeholders

ACS data: Many data users requested tools and infographics to help work with American Community Survey data, which "are really quite messy" and have more changes year to year than expected. (Community Solutions, International Downtown Association, National Healthy Start Association.)

Specific guidelines would also be helpful, e.g., using the 5-year ACS covering the COVID years versus the non-COVID years.

"We use the American Community Survey to develop our program community profiles. I love the ACS, but working with the ACS requires a lot of work. The numbers don't tell the full story. We still must add the context and connect the dots. Data stories and visualizations would be useful to communicate the full story to the community and write grant proposals." (National Healthy Start Association).

---

[2] https://downtown.org/the-value-of-downtowns/

Increase ACS sample size in disaster areas: Attendees advocated for an increase in the Census Bureau's budget to expand sample sizes and enhance data collection, particularly to monitor the migration and mobility of households in the aftermath of climate events like hurricanes and wildfires.

American Factfinder.

American Factfinder. Many data users at Listening Sessions noted that the new data tools (e.g., data.census.gov) are much harder to use than American Factfinder. (Community Solutions, International Downtown Association, National Healthy Start Association, NLIHC DHRRC.)

Time series, especially ACS: Creating tools to facilitate creating time series from cross-sectional household surveys. This request primarily focused on the American Community Survey. Every year, there are changes in the ACS requiring users to change their code. More sophisticated users have programs they adjust. Less sophisticated users often do this work using Excel or other software, requiring hours of work. Both expressed frustration in having to discover and make the changes. Similar changes to the API would also be helpful. (International Downtown Association, NLIHC DHRRC, Center for Risk-Based Community Resilience Planning.)

Integrate data from CDC, ACS, and other sources. Can tools be developed to integrate CDC health data and decennial and ACS demographic data to overcome the complexity of merging these data sets manually? Modeled or synthetic data would be acceptable to achieve granularity on CDC information, particularly for small area estimates. *Additional Stratifications*: To assess health disparities, users need health and demographic data of students by school enrollment by race and ethnicity for preschool, K through 8, and high school. (Community Solutions.)

Create baseline data products so that data users do not have to reinvent the wheel, using Census Bureau, local, and third-party data, such as housing and property data, property insurance data, social vulnerability indexes, credit scores, debt history, and third-party commercial data to score or assign risk as another way to measure high need communities. (Community Solutions, International Downtown Association, Third-party Data, NLIHC DHRRC.)

Standardize **social vulnerability indexes** or create several indexes and explain the differences. (Community Solutions, Center for Risk-Based Community Resilience Planning, NLIHC DHRRC.)

Administrative records and third-party data: The integration of government and third-party data sources with traditional Census Bureau surveys is seen as vital for enriching the statistical data ecosystem. This approach can enhance the granularity, timeliness, and relevance of data products. While there's interest in integrating third-party data, including satellite imagery, there's also a call for transparency and caution about relying on proprietary tools. This highlights the need for a balanced approach that maximizes data utility while ensuring data integrity and transparency. (Local Data Partnerships, Third-Party Data.)

Survey data: Traditional survey data remains a cornerstone. There is a push toward making these data collections more efficient and responsive to emerging trends and needs. The use of rapid response Pulse surveys for quick turnaround data was mentioned as a valuable tool, especially in capturing the dynamics of sudden societal and economic shifts, as seen during the pandemic. (Local Data Partnerships, Third Party Data.)

ACS questions about gender and sexuality: These questions are going in the right direction. The write-in options are a best practice for inclusivity if the multiple-choice answers do not fit the person. (Community Solutions.)

- Why is the cut-off age set at 15? This feels arbitrary.
- Does the person filling out the ACS survey on behalf of their household know the sexual orientation or the gender identity of everyone in their household? Many people do not feel safe reporting on their sexual and gender identity. Building trust is important.
- The Trevor Project (https://www.thetrevorproject.org/) provides data on LGBTQ youth statistics. Pew Research Center provides data on adults. (https://www.pewresearch.org/topic/gender-lgbtq/ .)

Community-Based Data: The potential for leveraging local data to supplement traditional Census Bureau sources was discussed, indicating a need for more flexible data collection policies and improved infrastructure at the community level to capture more accurate counts to understand the needs of underserved populations. (Healthy Start Association, Community Solutions, International Downtown Association.)

School Enrollment Data: Enrollment data could be a proxy for understanding the numbers of children in a specific area, which could help adjust for undercounts in decennial census data. (State Longitudinal Data Systems (SLDS). (Community Health Advocates, others.)

Internal Revenue Service Data: Useful for augmenting or supplementing household wealth characteristics missing from the decennial census. Specifically, IRS data could enhance understanding of health insurance marketplace enrollment, including the Affordable Care Act (ACA) marketplace credit and age stratifications. Tax Return Data: robust, little room for people to lie. (Community Health Advocates, Third-Party Data.)

Demographic data stratification needs to be more granular regarding types and geography. For example, the Asian category is too broad, as there are differences when studying Vietnamese, Korean, etc. In many school districts, students speak 30 to 90 different languages. (Community Solutions, others.)

Language Data: Data on spoken languages are needed to understand diverse community needs, particularly for hospitals and health departments to meet regulatory requirements. Current data are outdated and lack granularity at desired levels (e.g., block and block group levels, ZIP and sub-ZIP levels). (Health Advocates)

Ask supplemental survey questions: For example, Is gentrification an issue for your neighborhood? (Health Advocates)

Custom Geographies Data Wish List – Ideas from Listening Session stakeholders

Creating ACS geographies for downtown areas. IDA defines downtown as a city or town's central business district or main part. Downtowns can be grouped into one of three tiers based on their development stage, density, and significance to the city: emerging, growing, and established. These tiers are defined by the district's development stage both in terms of density and its significance to its city, as well as the growth of the district. (International Downtown Association, 2024)

Many asked for **block level data** so they could develop data around communities of interest, e.g., those hit by a hurricane or other climate event. In most cases, the request was to design or specify

geographic areas of interest, recognizing the value of synthetic or modeled data. (Community Solutions, NLIHC DHRRC, Downtown.org)

Create trend data with the challenges worked out. Many users want to use ACS data over time to assess changes in income, housing, and other topics collected in the ACS. However, this is challenging because every year, there are some variables that change how they are coded, and the geographies change each decade. Users recommended that the Census Bureau create and update annual ACS trend. This would save data users a lot of time and would ensure that the Census Bureau trend data are correct (Community Solutions, Healthy Start. NLIHC DHRRC, Downtown.org.)

Allow for **custom geographies** by neighborhoods and wards in the city. (Community Solutions has developed a methodology in collaboration with Cleveland State University.)

Discover, provide, and maintain uniform coding of data to facilitate comparisons across state: Housing valuations, building codes for each municipality, permit by types, roof-level data to show the condition of the home, coastal and riverine flooding data, loans and other funding to survive transitions from a climate event to finding stable housing, employment, etc. Potential data sources: Army Corps of Engineers, National Structure Inventory[3] (NSI). (NLIHC DHRRC)

Community Disaster **Resilience** Zones (CDRZs): Provide Social Vulnerability Indexes and resilience metrics data by CDRZs. (Center for Risk-Based Community Resilience Planning.).[4]

Targeted Pulse Surveys: Pulse surveys are valuable and were instrumental during the COVID-19 pandemic. Could they be used to gather information from areas affected by disasters, such as Florida after hurricanes? (NLIHC DHRRC.)

Health Data Wish List: Ideas from Listening Session stakeholders

Health data are needed to understand health trends across age groups and potentially to estimate population counts where Census Bureau data may be lacking, e.g., young children and underserved populations. (Health Advocates, Community Solutions.)

The Office of the National Coordinator has released new interoperability standards for electronic medical records (EMRs). Can social conditions and risks be more comprehensively connected?
- The new standards released by the Office of the National Coordinator are for interoperability for EMRs, specifically HL7 FHIR (Fast Health Interoperability Resources) standards. The idea proposed involves creating an underlying architecture or resource library associated with Census Bureau data that could map to HL7 FHIR resources.[5] This would facilitate communication between medical resources used in medical settings and community-based organizations, comprehensively contributing to a better understanding of social conditions and risks. The aim is to use HL7 FHIR standards to improve the integration of medical data with Census Bureau data, thereby enhancing the ability to effectively understand and address social determinants of health.
- For example, an individual has a medical appointment. The patient has conditions that have been identified. Would providing community-level data in a medical context be

---

[3] https://hydrologicengineeringcenter.github.io/NSI/
[4] . https://www.fema.gov/partnerships/community-disaster-resilience-zones
[5] https://www.hl7.org/fhir/overview.html

useful and powerful? Other initiatives funded by Medicare, Medicaid, and others are funding community health initiatives, recognizing that the community's health affects its residents' health. This is an example of where synthetic data would be important in providing data for small areas.

IRS Section 501r reporting: A hospital organization uses Internal Revenue Service Form 990, Schedule H, to provide information on the activities, policies, and community benefit provided by its hospital facilities and other non-hospital health care facilities operated during the tax year. A specific requirement includes doing a language assessment of hospital patients, which may require modeling to estimate. (Community Solutions.)

Small Area Estimates and Synthetic Data Wish List: Ideas from Listening Session stakeholders

While their methods are different, the use of synthetic data or small area estimates provides more insights into small populations at low levels of geography than are currently available in publicly available data. These data have the potential to allow users to create geographies of interest. Stakeholders would like access to small area estimates for rural, tribal, and other small areas to be defined by them, such as defining communities or communities in one part of a city. While the current Census Bureau products and tools are useful, they would like additional custom geographies at lower levels than the Census Bureau currently produces in SAIPE, SAHIE, LEHD and related data products and across a broader array of topics. There was strong interest in having access to synthetic data by many groups and tools to access and work with synthetic data. Here are some issues and ideas heard.

Organizations can only access public use files, so their models and estimates are not as detailed as Census Bureau-created small area estimates for small populations.

The use of synthetic data is sometimes difficult to explain to policymakers. Can more information be provided to help with this?

We need more detailed data for geographies that we define, such as specific areas of a city or town, rural areas, and unincorporated areas (that do not have building codes or other regulations). (Third Party Data, NLIHC DHRRC, Community Solutions)

Combining data from surveys to enhance usefulness requires model-based solutions, e.g., combine ACS and CDC Behavioral Risk Factor Surveillance System (BRFSS) data.

See discussion under **Health Data Wish List** about new interoperability standards for electronic medical records (EMRs) and combining those data with synthetic information about a patient's community.

Tools
Participants identified a variety of data tools that would facilitate their reporting for preparing grants and required for federal and state programs, and other work. These tools would help address their reporting needs and build their capacity for data sources and uses.

Tools to Address Reporting Needs and Build Capacity – Ideas from Listening Session stakeholders

Community Health Assessments (CHA): Over 5,000 hospitals and every public health department seek quantitative and qualitative information from their residents through town halls,

focus groups, and surveys to create a CHA, collecting data in various ways. Could the Census Bureau use machine learning techniques to process and synthesize the qualitative data? Would the survey data be useful to the Census Bureau and how? (Health advocates, Community Solutions.)

Head Start Performance Standards Data Set: Initiated in Minnesota to aid Head Start agencies across the state in meeting their reporting requirements using Census Bureau data. This interest highlights the need for targeted, usable tools and data products to simplify and enhance reporting for specific programs.

Federal Healthy Start Project Reporting: Organizations report various health-related measures to Health Resources and Services Administration, including demographic information, infant mortality rates, prenatal care, tobacco use, depression referrals, Inactivated Polio Vaccine screenings, breastfeeding rates, safe sleep practices, etc. These data are collected at the individual level by each Healthy Start organization and aggregated for reporting purposes. (National Healthy Start Association.)

Direct Data Sharing and Analysis by Local Entities and States: The potential for states to directly bring Census Bureau data into their systems, analyze it, and share results back was mentioned as a less legally complex model for collaboration, suggesting a flexible approach to integrating local insights into federal demographic analyses.

Data on Public Services and Infrastructure: A tool that addresses this question can inform strategies to bridge digital divides and promote inclusive growth. (Local Data Partnerships, Community Solutions)

Predictive Analytics and Scenario Modeling: There is a demand for data products that support predictive analysis and scenario planning, enabling users to explore "if, then" scenarios. This is particularly relevant for preparing for future events, such as climate-related disasters, and planning resource allocation or emergency responses. (Local Data Partnerships, others.)

Provide tools to build **administrative capacity** for local communities, especially small and rural communities. Need this for pre-disaster, not just post-disaster. (NLIHC DHRRC, Center for Risk-Based Community Resilience Planning)

Replicate the CDC Social Vulnerability Indexes. Allow users to add or delete variables. (NLHIC DHRRC, Center for Risk-Based Community Resilience Planning)

Tracking new Census Bureau products: Can the Census Bureau enhance its systems to better track and communicate new developments in data collection? Can they make it easier to track "What's New?"

Infographics
Every survey and question has a history, purpose, and population being observed. A first step is helping users understand the differences and why they might use data from one survey over another. One theme is identifying differences between household (demographic) and economic (business) surveys using infographics. These infographics could also provide a baseline for developing artificial intelligence (AI) tools that help users discover and choose data sources that meet their needs. Good infographics are an essential input to creating baseline datasets for statistical product development.

Infographics – Ideas from Listening Session stakeholders

Identify differences in definitions and measures across surveys and agencies so that users have a better understanding of the breadth of data sources, for example:

Occupation and labor force data are available from the Bureau of Labor Statistics (BLS) and Census Bureau surveys, such as the Quarterly Census of Employment and Wages (QCEW), the American Community Survey (ACS), the Current Population Survey (CPS), and others. For example, an area of policy interest is understanding manufacturing employment across industries (NAICS) and sectors. (Expert Interviews.)

Housing inventory and shortages. To measure housing shortages more accurately, better alignment between household formation, housing inventory, and building permits data is needed. (Expert Interviews.)

Social Vulnerability Indexes from CDC and others (Center for Risk-Based Community Resilience Planning.)

Resilience and other measures – There are dozens of resiliency metrics. Provide a description and use for each one. There are varying definitions of related terms such as gentrification, highlighting the complexity and potential misinterpretations in data related to events such as Hurricane Helene's impact on Asheville residents. (NLIHC DHRRC, Center for Risk-Based Community Resilience Planning.)

Guidelines on when to use one-year and five-year ACS data. What are the differences between the two? Why are the estimates different? What should be used for population projections? Provide *What If* scenarios. (Center for Risk-Based Community Resilience Planning.)

Guidelines on how to conduct local surveys. (Community Solutions.)

Eliciting Information About Purposes and Uses – Ideas from Listening Session Stakeholders

Ensuring Participation in Seeking Stakeholder Ideas.
- Implement a feedback mechanism on data utilization platforms to understand user intentions and requirements.
- Work intentionally with community-based organizations with deep connections to the communities of interest to ensure diverse voices are heard.
- Validate findings and data interpretations with the communities they represent to ensure accuracy and relevance.
- Establish survey guidelines to ensure they capture the necessary information for the Census Bureau and other stakeholders in a manner that is inclusive and representative of all community segments.

Building trust and focusing on strengths in communities.
- "We learned that our data **presentations were problem and deficit-focused instead of strength-based.** This approach is not good for trust-building. We need to focus on opportunities." (Community Solutions.)
  - Small sample sizes make it challenging to work at the community level.

- One issue is that synthetic data are sometimes difficult to explain to policymakers. (Note: Participants used the term synthetic data to mean data created from published and microdata to create data for small areas. They were not specific about the approach to use.)
- It is critical to involve the community in decisions about changes to the survey.

Census Bureau Partnerships- How do we Collaborate with the US Census Bureau?

Census Bureau vetting or utilizing data as a mark of quality. This endorsement could significantly enhance the credibility and trust in data provided by third parties, making it a valuable benchmark for analytics and decision-making in various sectors. (Third Party Data.)
- Ability to benchmark.
- Ability to create a synthetic dataset based on consumer information (Census Bureau) and credit card records of purchases (private data) to understand consumer demands better.
- Monetizing data results.
- Federal agencies can bring together expertise in areas that private industry may also be interested in (i.e., supply chains).
- Consortium database for statistical purposes.
- User understanding of curated and authoritative data.

Administrative data-sharing with state agencies and the private sector. There is potential for collaboration between state agencies and the Census Bureau to share administrative data, highlighting the value of administrative records in enhancing Census Bureau data and other statistical products. (Early Childhood Integrated Data Systems, NLIHC DHRRC). For example:
- Taking advantage of state initiatives focused on integrating children's participation in state and local programs focused on health, well-being, and education.
- Engagement with State Housing Finance Agencies, which possess relevant data and are typically willing to share information and support research efforts.
- Collaboration with insurance companies: Participants emphasized the importance of partnering with insurance companies to access their extensive data, alongside leveraging additional data sources, such as the extensive third-party data being acquired by the Census Bureau.

Participants encouraged the Census Bureau to re-imagine its approach to data dissemination, particularly in providing tailored support and resources to organizations with varying levels of data literacy and capacity. Ideas proposed:
- **Concierge services**. Participants expressed interest in engaging more directly with Census Bureau efforts, specifically mentioning the potential for collaboration with the Young Children Working Group.
- **Online Chat**. The need for technical support through an online chat option for real-time help, was mentioned as a practical way to support community-based organizations in utilizing data more effectively. This could be one part of an overall concierge service.

Listening Sessions at Conferences

American Community Survey Data Users Conference Listening Session, May 18, 2023

After short presentations describing the history and approach to developing the Curated Data Enterprise, we discussed the audience's ideas for purposes and uses. Because of the large audience, we also interspersed the discussion with polling questions.

As the data show, many attendees (57%) said that more geographic detail would add the most value to their statistical products. More than two-thirds use social and demographic data when applying for grants. Almost 40% suggested eliciting information about purposes and uses by increasing outreach and education. The other approaches were each about 20%: recommended convening listening sessions, establishing working groups, and inviting proposals from users.

What would add the *most* value to your statistical products? (n=67)	
Geographic detail	57%
Specific topics (e.g., housing)	21%
Timelier estimates (<1 year old)	22%

Do you need to use social and demographic data when applying for grants (N=46)	
No	30%
Yes	70%

How do we elicit the information about purpose and use?	
Convene listening sessions	18%
Establish working groups	20%
Increase outreach & education	39%
Invite proposals from users	23%

In the response to the open-ended question, we categorized the comments by keyword and provided selected excerpts from the open-ended questions:
- Training: e.g., ACS webinars are super helpful and are a great resource. Sessions like this are awesome to be able to communicate with you all.
- Citations: I still don't know how to cite Census Bureau data in a publication (See US Census Bureau, "Citing our Data, Tools, Technical Documents, and Research."[6]
- Software interface: The API is great! I mainly access it via {tidycensus}.[7] Increased functionality and collaboration with external researchers aiming to make data accessible are essential.
- Data Formatting: To meet data users where they are, always have a simple CSV download of any dataset. There were also requests for data in Spanish, Braille, and other formats.

---

[6] https://www.census.gov/about/policies/citation.html#:~:text=U.S.%20Census%20Bureau%2C%20%E2%80%9CTitle%20of%20Webpage%E2%80%9D%20date%20of%20posting,accessed%20on%20August%2027%2C%202024

[7] The R software package {tidycensus} that allows users to interface with the US Census Bureau's decennial Census and five-year American Community APIs and return tidyverse-ready data frames.

- Customization: Customize cross-tabulations for areas smaller than Public Use Microdata Sample files; Add an option to select a table layout that highlights cells in the tables with high margins of error (e.g., MOE GTE 50%) to provide a quick way to see less reliable values instead of including the columns that need interpretation.
- Classifications and subgroups: It would be helpful to enable real-time collapsing of selected subgroups in ACS tables. It is onerous to download and manually do this offline.
- Group quarters (GQ): GQ is extremely important because we have so many college students across the state! (Boston, Cambridge, Amherst, etc.)
- Local level data: Neighborhood-based data; General well-being (or lack thereof) of our residents.
- Organizational collaboration: More cross-over with the Bureau of Labor Statistics, CDC, other agencies. Work with state data centers to form relationships with smaller government organizations (e.g., municipalities) and non-profit organizations. Most smaller entities may not have the capacity to stay up to date with what's happening if it's not in front of them.
- Race/ethnicity data: My users would need more detail about race/ethnicity, immigration from other countries, income/poverty, and housing (tenure, cost burden). Other groups need to be defined in more detail as well. These include Hispanic, Middle East and North Africa, separate Asian categories.
- Rural, Tribal, School Districts: How can rural areas be better represented? More data for school districts from all Census Bureau data sources. Data for Tribal lands single entries (climate change, P&C).

Association for Public Data Users Annual Conference – a Listening Session led by Steve Jost, July 26, 2023. There were about 60 participants.
- Encouraged the Census Bureau to tackle legal and statutory barriers to using administrative data.
- Suggested we work on uniform race standards across federal agencies and applauded the Census Bureau's work to address the undercount of children for 2030.
- Ensure privacy and confidentiality.
- Encouraged the Census Bureau to seek purposes and uses from underserved and hard-to-reach populations by getting out of DC. (Other attendees echoed this sentiment.)
- There were many positive statements about the CDE like, *this is exciting*, *very encouraging*, and there was a lot of positive head-nodding throughout the session.

Prioritizing Purposes and Uses.

This section discusses one approach to prioritizing purposes and uses.

Listening Session Follow-up.

The National Low Income Housing Consortium (NLIHC) Disaster Housing Recovery Research Consortium (DHRRC)[8] invited us to follow up on our May Listening Session. At the May session, we[9] introduced the CDE and broke into four groups (2 with researchers and 2 with advocates) to identify their purposes and uses and data wish lists. At the October meeting, we[10] sought input on prioritizing the purposes and uses identified in the May Listening Sessions and discussions with experts.

We asked the participants to use a 2x2 impact-effort matrix to prioritize Purposes and Uses. Using this approach, Purposes and Uses are assigned to a quadrant based on participants' rating of their potential impact and manageable effort. It quickly became evident that this framework was challenging. It is hard for people to keep four options in their heads, e.g., low-low, low-high, high-low, high-high (see Exhibit 3).

Exhibit 3: A test to assess if a 2x2 matrix is useful to prioritize Purposes and Uses.

One suggestion for a more effective approach is to use a 4-part question that enables better comprehension and categorization of options. The phrasing is slightly different but works better as a 4-part question. The answers can then be presented in a 2x2 matrix. (See the discussion below about using tools to help with the prioritization.)

---

[8] The NLIHC-led Disaster Housing Recovery Research Consortium Coalition (DHRRC) works with Congress, FEMA, and HUD "to ensure that federal disaster recovery resources reach all impacted households, including those with the lowest incomes who are often the hardest hit by disasters and have the fewest resources to recover afterward." https://nlihc.org/disaster-housing-recovery-research-resilience

[9] "We" included Stephanie Shipp and US Census Bureau colleagues Robin Bachman, Bethany DeSalvo, Chip Walker, and Karla Riso each of whom led one of the four Listening Session breakout groups.

[10] "We" included Stephanie Shipp and US Census Bureau colleague, Erika Becker-Medina.

The participants discussed characteristics for prioritizing Purposes and Uses focused on fairness and resilience:

- Relevance: How well does the Purpose and Use align with current issues, needs, or emerging trends in the field?
- Novelty: Does the Purpose and Use address a new question or offer a unique perspective on an existing issue?
- Significance: What is the potential impact of the findings on knowledge, practice, or policy?
- Feasibility: Is the Purpose and Use feasible considering available resources, timeframes, and technical capabilities?
- Ethical Considerations: Are there potential ethical concerns related to the Purpose and Use?

Use of Group Concept Mapping to Prioritize Purposes and Uses.

Group concept mapping is a structured methodology for organizing and visually representing a group's ideas on a specific topic through interrelated maps. Developed by William M.K. Trochim in the late 1980s at Cornell University, it combines qualitative and quantitative approaches, allowing diverse participants to express their thoughts collaboratively (Kane and Trochim 2007). The process involves brainstorming, sorting, and rating ideas, followed by data analysis using specialized software to create maps that illustrate the relationships between concepts. This participatory approach has been applied across various fields, including community health, social work, and biomedical research, helping groups to develop conceptual frameworks that guide evaluation and planning efforts.

Group Concept Mapping (GCM) Resources.

Tulane University's Disaster Resilience Leadership Academy uses Group Concept Mapping (GCM). A member of the NLIHC DRHC group and a member of the Tulane's Academy, Tim Davidson, provided examples and references to facilitate the use of Group Concept Mapping (Ferreira & Davidson 2024).

- A GCM Research brief: A digestible and explanatory application (Ferreira & Davidson 2024).
- GCM database of published articles: look here for example articles that have been published to get more ideas of what this looks like on paper.
- GCM resource guide: The website provides a description of GCM and why to use this approach, GCM steps: The steps describe how to use the groupwisdom™ tool, training opportunities, and templates:
  - Blank data cleaning template: Use this template to clean the data.
  - Data cleaning example: The examples provide selected cells to see how to do the cleaning.

The steps used in the groupwisdom™ tool include recruitment and participation (e.g., Listening Sessions), sorting and rating the participant statements, and analysis. Using this tool, the GCM process involves both qualitative and quantitative methods (Ferreira & Davidson 2024):

Preparation: This initial step includes developing focus prompts and recruiting experts for the study. It sets the groundwork for the entire project.

Generating Ideas: Participants brainstormed in group to generate statements related to the study's focus.

Sorting: Participants sort the generated statements into piles based on their conceptual similarities. This activity helps to organize ideas meaningfully. The output is a similarity matrix that indicates how frequently pairs of statements were sorted together.

Developing Maps: A point cluster map is created using the sorted data to visualize relationships among the statements. Multidimensional scaling (MDS) uses the similarity matrix to plot each statement onto a two-dimensional plane where statements that appear frequently together appear closer to each other.

Hierarchical Cluster Analysis: This step uses the MDS (created in prior step) and suggests ways to group the statements based on their proximity on the map. The optimal number of clusters was determined based on conceptual coherence and interpretability.

Interpreting Results: This step involves analyzing the point map through hierarchical cluster analysis to identify distinct themes from the clusters, enhancing understanding of the data. The naming of clusters is done by reviewing the statements within each cluster and considering participant-suggested labels and overarching themes.

Rating Analysis: Participants rate each statement on scales of terms relevant to the study. In this case, "equity and resilience" were the terms used, which is very similar to the work in this report. These ratings are used to identify "go-zones," which are areas on the cluster map where statements are rated highly on both scales – the top right quadrant (see Exhibit 4).

Exhibit 4. Rating Analysis: Example Go-Zone chart from CEDR's group concept mapping study exploring barriers to recovery showing average importance and feasibility to address. (Source: Ferreira & Davidson 2024)

Pattern match analysis: Compares the average cluster ratings of two scales, illustrating which clusters were more critical to resilience. Pattern match maps provide a visual comparison of different themes, helping researchers see which clusters are most representative of resilience and aid in interpreting the results. These are portrayed in the pattern match (Exhibit 5) and cluster maps (Exhibit 6).

In conclusion, group concept mapping emerges as a methodology for organizing and visually representing a collective set of ideas on a given topic. The structured process includes brainstorming, sorting, and rating ideas, followed by data analysis with specialized software to create maps that elucidate the relationships between concepts. Its participatory nature facilitates the development of conceptual frameworks that harness the collective wisdom of stakeholders, leading to more informed decision-making and prioritization of ideas.

Exhibit 5: Criteria Cluster Pattern Match (Source: Ferreira & Davidson 2024).

Exhibit 6. Cluster maps are Top Down and Bottom Up (Source: Ferreira & Davidson 2024).

Task 2: Developing a Curated Data Enterprise (CDE) Research Agenda.

Use Case Research Program to Develop a Research Agenda.

Many use cases will be needed to define the capabilities to build in the CDE. Examples of these capabilities include algorithms, data integration, statistical and data science methods, visualization tools, dissemination mechanisms, and a curation pipeline. These capabilities need to be developed in the context of a Use Case to make them tangible. These capabilities form the foundation for creating a research agenda.

Use cases are a powerful mechanism to engage stakeholders and the research community and promote methodological research to develop and implement the capabilities needed to deploy the proposed CDE infrastructure successfully. The objectives are to engage a broad array of researchers capable of undertaking research projects that have the potential to create CDE capabilities and who will embrace the End-to-End (E2E) curation components. These capabilities and curation outputs can be used to create statistical products that address stakeholders' Purposes and Uses.

We define the process of identifying and implementing use cases as a research approach to identify, clarify, and organize requirements to create a CDE and end-to-end curation pipeline. A Use Case provides context that defines a process or processes to formulate a research question, discover, acquire, curate, and store data; develop a hierarchy for data sources on the same topic; benchmark or validate synthetic data; and develop methods for creating statistics. These capabilities will populate the CDE and make possible statistical product production in new ways. The CDE Framework provides the roadmap to guide this work (see Exhibit 1).

Art-of-the-possible use cases are focused on current policy issues for which there is no single or definitive data source to answer policy questions. These entail a literature review and data discovery to assess if a demonstration use case is warranted. When fully implemented, these proposed use cases demonstrate a collection of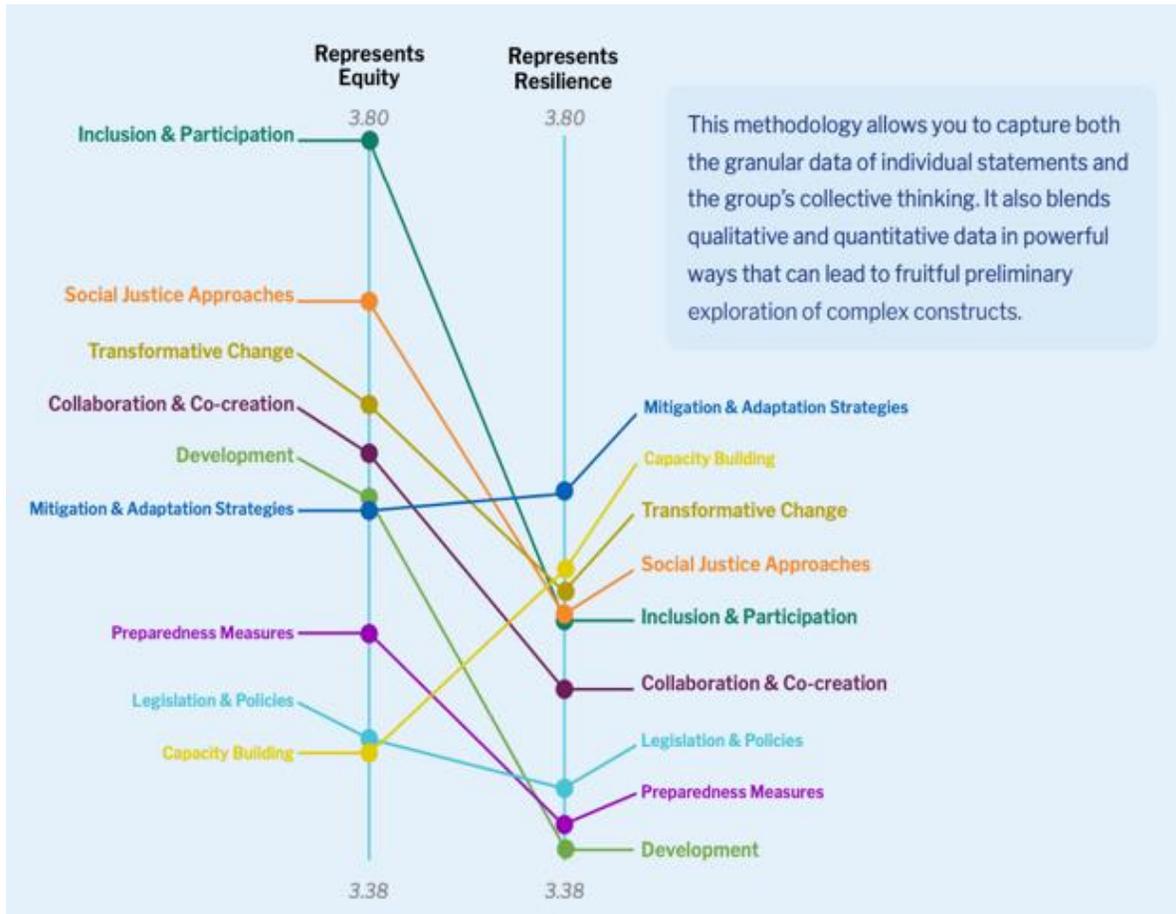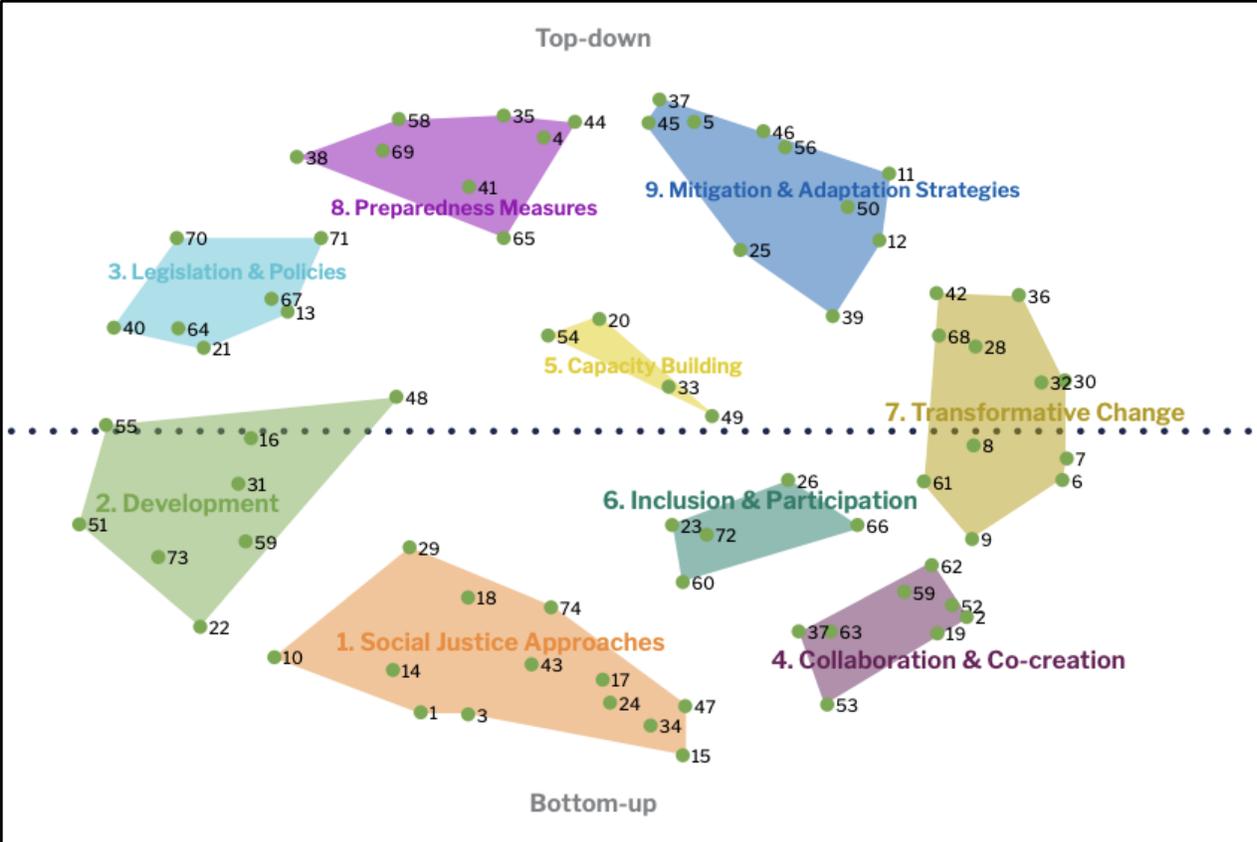 capabilities that need to be built into the CDE, such as reusing modules and data, tracking the provenance of collected and generated data, creating and reusing synthetic data and methods to integrate many types of data, conducting statistical analyses, and reviewing data and results through a fairness and ethics lens.

We defined the Use Case Research steps using the CDE framework as a guide (see Exhibit 1). Using that as our starting point, we created **Use Case selection and evaluation criteria** (described in the next section) and socialized them through discussions with researchers and others as the foundation for a Use Case Research Program Design.

To better understand the development of Uses Cases, we piloted the criteria in a demonstration Use Case to create a Household Living Budget (HLB) at the Census Bureau tract level (Lancaster et al. 2023 and 2024). Then, we applied the HLB to food insecurity in a second demonstration (Montalvo et al. 2023 and 2024) (see Appendix 3). We further refined the criteria by testing them against an earlier Use Case on the resiliency of skilled nursing facilities during an extreme climate event (Lancaster et al. 2023). See Appendix 3 for summaries of these demonstration use cases.

We completed two art-of-the-possible use cases to assess their feasibility to be a demonstration Use Case and to assess the Use Case selection criteria (see Appendix 3). The first was to assess the feasibility of a demonstration study to address undercount of young children, highlighting state-based Early Childhood Integrated Data Systems (Salvo, Lancaster & Shipp 2023). The second was to identify the issues related to the housing affordability crisis for low-income households and discover data sources (Wu et al., 2023). These are summarized in Appendix 3.

We also completed a scan of State Data Centers and related state activities (Goswami et al. 2024). See Appendix 3 for a summary. We described how selected international statistical agencies work to integrate administrative data into their statistics (Lanman, Davis, & Shipp, 2023). See Appendix 7.

Use Case Selection and Evaluation Criteria

We developed Use Case selection and evaluation criteria to guide and prioritize the selection of Use Case research projects. We established initial criteria, vetted them with experts and listening session participants, and piloted them with a Use Case on creating a new Household Living Budget with an application to food insecurity. We tested the criteria using an earlier developed Use Case for measuring the resiliency of Skilled Nursing Facilities during an extreme climate event (Lancaster et al. 2023). We refined the criteria based on what was learned from these activities.

<u>**Criteria for Selecting and Evaluating Use Cases to Identify CDE Capabilities.**</u>

Value and feasibility of the CDE approach described in the existing research (potential Use Case case) to address emerging, long-standing, and high-priority issues, i.e., purposes and uses over and above existing approaches.

Stakeholders' challenges and issues as the source of purposes and uses.

Subject matter experts to advise on the content, approach, and implementation.

Partners to access data from local and state governments, nonprofit organizations, and the private sector and to develop strategies to overcome the legal and administrative barriers to access that benefit both the providers and recipients of the data.

Designed, administrative, opportunity, and procedural data from multiple sources (e.g., local, state, federal, third-party) to address the purposes and uses (issue) in an integrated way through well-defined data ingestion and governance requirements.

Computation and measurement requirements for statistical products such as the unit(s) of analysis and their characteristics, temporal sequence, geographic detail, imputation methods, projections, synthetic populations, and statistical analyses.

Fairness and ethical dimensions considered at each step to ensure the Use Case provides accurate representation across groups and an assessment that the potential benefits outweigh the potential drawbacks.

Evidence of CDE capabilities to be built, including the code, data, metadata, paradata, and documentation to create the statistical products, which can be described through the curation steps.

Statistical products include integrated data sources, composite indicators, maps, visualizations, storytelling, and analyses.

Potential viability of proposed **dissemination platforms** for interactive access to integrated data products at all levels of data acumen (Keller & Shipp 2021) while adhering to confidentiality and privacy rules.

Overview of Key Components of a Use Case Research Program.
The CDE Use Case Program is designed to support the rapid development of statistical products. There are four parts to the program.

1. Criteria for inviting/identifying, selecting, deploying, and evaluating use cases that will uncover the capabilities and research necessary to develop the CDE.
2. An End-to-End (E2E) Curation process to ensure reusability, transparency, and repeatability (Nusser 2023).
3. Selection and invitation of researchers to participate in the CDE Use Case program.
4. Translating and creating a roadmap for the Census Bureau to build CDE capabilities through synthesizing the CDE Use Case Program outcomes.

When implemented, these proposed use cases will demonstrate a collection of capabilities that need to be included in the CDE, such as agile data discovery, reusing modules and data (including synthetic data), tracking the provenance of collected and generated data, reusing synthetic data, methods to wrangle many types of data, conducting statistical analyses involving different analysis units, processes to ensure ethics and fairness from the elicitation of purposes and uses to the dissemination of statistics, and end-to-end curation.

Our discussions with researchers, including Federal Statistical Research Data Centers (FSRDC) researchers, sought their input on the CDE Use Case Program design and the application of the CDE framework as a guide to curating their research. They were enthusiastic about the CDE approach and the promise of creating statistical products that are timelier and more comprehensively address social and economic issues. (See Appendix 4 for a summary of our discussions with FSRDC researchers.)

We also asked them what their motivations would be if asked to participate in curating their research end to end. Overall, there were positive reactions to the CDE framework and emphasis on curating the entire statistical product development process. Some researchers said they essentially use a similar approach as the CDE framework to guide and document their research. Others said that with some modest support they could curate their research using the CDE framework.

C. Findings and Conclusions

The organization and execution of numerous listening sessions yielded valuable insights from participants regarding their unanswered questions related to existing data (Purposes and Uses). These sessions highlighted participants' requests to create infrastructural data products, infographics, data tools, and opportunities for collaboration with the US Census Bureau.

Engaging stakeholders emerged as a crucial step for further advancing the research agenda, gathering insights, and establishing high-level requirements for the Curated Data Enterprise (CDE). A recommended approach is to develop a prioritization strategy, such as the Global Concept Mapping tool described above, to prioritize Purposes and Uses. This is one example of a strategy to guide the selection of future use cases and Statistical Products.

Two types of use cases assessed the identification and creation of CDE capabilities. We completed two Art of the Possible Use Cases on the undercount of young children and housing affordability. In these use cases, we obtained expert input through literature reviews and discussions, exploring the feasibility of various data sources to address the purposes and uses. Additionally, we conducted demonstration use cases to create a Household Living Budget (HLB) and food insecurity metrics. To develop the CDE, this work must be refined, and many new Use Cases developed.

The Census Bureau already collaborates with several scientific agencies that have extensive data sources, including **National Oceanic and Atmospheric Administration** (NOAA), National Institutes of Standards and Technology (NIST), Federal Emergency Management Agency (FEMA), Department of Housing and Urban Development (HUD), US Geological Survey (USGS), and US Department of Agriculture (USDA). Identifying potential partners, such as NOAA and NIST, and establishing connections with the National Low Income Housing Coalition can strengthen collaborative efforts. Engaging with the University of Pennsylvania-led Actionable Intelligence for Social Policy (AISP) network could provide valuable local data and opportunities for developing demonstration use cases and statistical products for specific regions. Given the focus on addressing the undercount of young children, potential partnerships with organizations like the Annie E. Casey Foundation could enhance the understanding of children's data, including through a demonstration use case utilizing State Early Childhood Integrated Data Systems (ECIDS) data across multiple states.

Use cases are generally created for specific geographic areas as prototypes, analogous to the Census Bureau's Experimental program. These use cases could be scaled to regional or national levels, allowing for testing of scalability and validation. Securing funding for the CDE Use Case Research Program is essential for developing a comprehensive library of use cases that enhance CDE capabilities. Additionally, the integration of AI and the dissemination of emerging indicators will be pivotal aspects of this initiative, alongside the need to develop requirements for an end-to-end Curation Pipeline.

Ultimately, developing CDE use cases will inspire stakeholders to address pressing societal issues and cultivate public understanding and trust in federal statistics. We anticipate that the successful implementation of the CDE research agenda will yield timely statistical products that address the Purposes and Uses asked by diverse communities, including tribal nations, rural areas, towns, inner cities, and coastal areas with large pockets of vulnerable populations.

References.

Reports delivered (* since July 2022, since the first Summary report**).**

*Goswami T, Cai J, van der Geer M. Lancaster V, Kattampallil N, Shipp S, Molfino E. **A Review of State Data Centers Indicators, Tools, and Methods Technical Report,** 2024 Proceedings of the Biocomplexity Institute, Technical Report. University of Virginia. https://doi.org/10.18130/hbnk-ah54.

*Lancaster V, Shipp S, Keller S, Schroeder A, Mortveit H, Swarup S, Xie D, (2023), **Census Curated Data Enterprise Use Case Demonstration: Climate Resiliency of Skilled Nursing Facilities.** Proceedings of the Biocomplexity Institute, Technical Report. TR# 2023-53. University of Virginia. https://doi.org/10.18130/ce97-sp05.

*Lancaster V, Montalvo C, Salvo J, Shipp S, **The Importance of Household Living Budget in the Context of Measuring Economic Vulnerability: A Census Curated Data Enterprise Use Case Demonstration** Proceedings of the Biocomplexity Institute, Technical Report. TR# BI-2023-258. University of Virginia; 2023 October. DOI: https://doi.org/10.18130/p43z-c742.

*Lanman K, Davis O, Stephanie S. **International Statistical Agencies: What can we learn from other countries about how they are using administrative data to supplement, enhance, or create new data products?** Technical Report BI-2023. Proceedings of the Biocomplexity Institute, University of Virginia; 2023 October. https://doi.org/10.18130/2n54-sc22.

*Montalvo C, Lancaster V, Salvo J, Shipp S, **The Importance of Household Living Budget in the Context of Food Insecurity: A Census Curated Data Enterprise Use Case** Demonstration (2023) Proceedings of the Biocomplexity Institute, Technical Report.TR BI-2023-261. University of Virginia. https://doi.org/10.18130/2kgx-tv50.

*Salvo J, Lancaster V, Shipp S. **The Net Undercount of Children Under 5 Years of Age in the Decennial Census:** An Art of the Possible Use Case. Proceedings of the Biocomplexity Institute, Technical Report. TR# BI-2023-000. University of Virginia; 2023 October. DOI: https://doi.org/10.18130/nzyj-m621.

Salvo J, Shipp S, Zhang S. **Defining the Role of Gig Employment in the Post-Pandemic World of Work**, Technical Report BI 2022-026. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. https://doi.org/10.18130/wkx0-4y46.

Salvo J, Shipp S, Zhang S. **Building a Case Study of Domestic Migration** and the Curated Data TR# 2022-027 – Essential Elements, Technical Report BI 2022-027. [Print]. Proceedings of the Biocomplexity Institute, University of Virginia; 2022 April. https://doi.org/10.18130/bcwa-gt69.

*Wu E, Salvo J, Lancaster V, Stephanie S. **Housing Affordability – An Art of the Possible Use Case to Develop the 21st Century Census Curated Data Enterprise**, Technical Report BI-2023-262. Proceedings of the Biocomplexity Institute, University of Virginia; 2023 October. DOI: https://doi.org/10.18130/qgkd-va29.

Journal Articles

Lancaster, V., Montalvo, C., Salvo, J., & Shipp, S. (2024). A New Measure of Economic Vulnerability for Better Policymaking – Creating a Household Living Budget Using Nonsurvey Data at the Census Tract Level, *Submitted to Statistics and Public Policy, September 2024.*

Lancaster V, Shipp S, Montavlo C, Salvo J, Shipp S, (2024). Creating a Household Living Budget Using Publicly Available Data: A 21st Century Census Curated Data Enterprise Use Case. CHANCE, 37(1), 22-32

https://www.tandfonline.com/eprint/HGCKEAP73RIDAAX4U9BU/full?target=10.1080/09332480.2024.2324664.

Montalvo, C., Lancaster, V., Salvo, J. J., & Shipp, S. (2024). A novel approach to assessing food insecurity for small geographic areas using household living budgets. Food Security, 1-18. https://doi.org/10.1007/s12571-024-01494-5.

Nusser, S. M. (2023). The role of statistics in promoting data reusability and research transparency. *Annual Review of Statistics and Its Application*, 10 (1), 145-164.

Published in Fall 2024, ***Real World Data Science***, a four-part article about the Curated Data Enterprise.

Keller S, Shipp S, Lancaster V, Salvo J. (2024) Part 1: Advancing Data Science in Official Statistics – The Policy Problem. Real World Data Science. https://realworlddatascience.net/case-studies/posts/2024/11/01/policy-problem.html.

Keller S, Shipp S, Lancaster V, Salvo J. (2024) Part 2: Advancing Data Science in Official Statistics – What is the Curated Data Enterprise. https://realworlddatascience.net/case-studies/posts/2024/11/08/what-is-CDE-2.html.

Lancaster V, Shipp S, Keller S, Mortveit H, Swarup S, Schroeder A, Xie D. (2024) Part 3: Translating the Curated Data Model into Practice: Climate resiliency of skilled nursing facilities, Real World Data Science. https://realworlddatascience.net/case-studies/posts/2024/11/19/use-case-2.html.

Shipp S, Lancaster V. Salvo J. (2024). Part 4: Defining Purposes and Uses to Support the Development of Statistical Products in a 21st Century Census Curated Data Enterprise Environment. Real World Data Science. https://realworlddatascience.net/case-studies/posts/2024/11/22/development-plan-2.html.

References
Ferreira, R. & Davidson, T. 2024. Group Concept Mapping and Equitable Resilience in Southern Africa. Research Summary Brief. Southern African Resilience Academy. https://www.globalresiliencepartnership.org/wp-content/uploads/2024/07/research-summary-brief_04july24.pdf.

Foster L. (2023) Evidence from Micro Data, Federal Statistics Research Data Center (FSRDC) Annual Research Conference. Federal Reserve Board, Washington DC. September 22.

Kane, M., & Trochim, W. (2007). *Concept mapping for planning and evaluation*. SAGE Publications, Inc. https://doi.org/10.4135/9781412983730.n1.

Keller S. (2024). Update on the Statistical Product First Approach. Census Scientific Advisory Committee Spring Meeting. March 14-15. https://www.census.gov/about/cac/sac/meetings/2024-03-meeting.html.

Keller S, Prewitt K, Thompson J, Jost S, Barrett C, Nusser S, Salvo J, Shipp S. (2022) A 21st Century Census Curated Data Enterprise. A Bold New Approach to Create Official Statistics, Technical Report BI-2022-1115. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. DOI: https://doi.org/10.18130/r174-yk24.

Keller S, Shipp S. (2021) Data Acumen in Action. American Mathematical Society. October. https://www.ams.org/journals/notices/202109/noti2353/noti2353.html?adat=October%202021&trk=2353&galt=feature&cat=feature&pdfissue=202109&pdffile=rnoti-p1468.pdf.

Keller S, Shipp S, Schroeder A, Korkmaz G. (2020). Doing Data Science: A Framework and Case Study. Harvard Data Science Review [Internet]. Feb 21;2(1). https://hdsr.mitpress.mit.edu/pub/hnptx6lq

National Academies of Sciences, Engineering, and Medicine: (2024). Principles and Practices for a Federal Statistical Agency: Eighth Edition. Washington, DC: The National Academies Press. Doi: 10.17226/27934.

National Academies of Sciences, Engineering, and Medicine. 2023. *Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26804.

National Academies of Sciences, Engineering, and Medicine. 2023. *An Updated Measure of Poverty: (Re)Drawing the Line.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26825.

National Academies of Sciences, Engineering, and Medicine. 2021. *Transparency in Statistical Information for the National Center for Science and Engineering Statistics and all Federal Statistical Agencies.* Washington, DC: The National Academies Press. https://doi.org/10.17226/26360.

Santos RL (2024). Looking Ahead to 2024 on Our Collective Journey to Excellence. Feb 26, 2024. https://www.census.gov/newsroom/blogs/director/2024/02/looking-ahead-to-2024.html#:~:text=As%20part%20of%20our%20transformation,continuous%20improvement%20and%20clear%20understandings.

Santos RL (2023). Blog: A Look Ahead: 2023. https://www.census.gov/newsroom/blogs/director/2023/01/a-look-ahead-2023.html.

Santos RL (2022). Blog: Thinking Differently About Perpetuating Excellence. https://www.census.gov/newsroom/blogs/director/2022/02/thinking-differently-perpetuating-excellence.html.

Thieme M. 2022. Technology Transformation at the Census Bureau: Building a Modern, Data-Centric Ecosystem October 5. US Census Bureau. https://www.census.gov/newsroom/blogs/research-matters/2022/10/technology-transformation.html.

US Census Bureau (2022). 2020 Census Data Quality Results: Post Enumeration Survey and Demographic Analysis, March 10. https://www.census.gov/content/dam/Census/newsroom/press-kits/2022/20220310-presentation-quality-news-conference.pdf.

US Office of Management and Budget (OMB). 2023. Fundamental Responsibilities of Recognized Statistical Agencies and Units. August. Federal Register. The Daily Journal of the US Government. https://www.federalregister.gov/documents/2023/08/18/2023-17664/fundamental-responsibilities-of-recognized-statistical-agencies-and-units.

Vilhuber L, Koren M, Llull J, Connolly M, Morrow P, (2022) A template README for social science replication packages. Social Science Data Editors, https://social-science-data-editors.github.io/template_README/.

Vilhuber L. 2023, Transparency and Reproducibility in Economics: Context and Lessons learned from 1,000 papers. https://zenodo.org/records/8362024.

Appendices

Appendix 1. Curated Data Enterprise (CDE) Team and US Census Bureau Advisors.

Appendix 2: Listening Sessions Summaries.

Appendix 3: Demonstration and Art of the Possible Use Case Summaries.

Appendix 4: Discussions with FSRDC Researchers and Experts.

Appendix 5. Stakeholder and Expert Meeting Summaries Around Resilience and Related Topics.

Appendix 6 Meeting with Other University Researchers and Agency Program Leaders.

Appendix 7: International Discussions.

Appendix 1. Curated Data Enterprise (CDE) Team and US Census Bureau Advisors

Our team included an experienced group of Census Bureau experts and leaders:
  Stephanie S. Shipp, Acting Director and Professor, Social and Decision Analytics Division,
       Biocomplexity Institute, University of Virginia, now at Iowa State University.
  Joseph Salvo, former Chief Demographer at NYC Department of City Planning, Institute Fellow,
       Biocomplexity Institute, University of Virginia.
  John Thompson, former US Census Bureau Director, 2013 to 2017, Distinguished Institute Fellow,
       Biocomplexity Institute, University of Virginia.
  Ken Prewitt, former US Census Bureau Director, 1998-2001, Carnegie Professor of Public Affairs and
       Special Advisor to the President, Columbia University.
  Stephen Jost, Senior Vice President, AVOQ; prior to the current position, oversaw US Census Bureau
       outreach and promotion campaigns for the 2000 and 2010 Census.
  Chris Barrett, Executive Director, Biocomplexity Institute, University of Virginia.
  Joan Naymark, Executive Director at MACS Minnesotans for the American Community Survey and
       2030 Census.
  C. Mathew Snipp, Burnet C. and Mildred Finley Wohlford Professor of Humanities and Sciences,
       Department of Sociology, Stanford University.
  Zhengyuan Zhu, Professor, Director of the Center for Survey Statistics and Methodology, Iowa State
       University.
  Erika Becker-Medina, US Census Bureau.

Academic Team
- Stephanie S. Shipp, Economist.
- Joseph Salvo, Demographer.
- Vicki Lancaster, Statistician
- Cesar Montalvo, Mathematician and economist.
- Edward Wu, Statistician.

Census Bureau Advisors.
1. Christa Jones, Chief of Staff, US Census Bureau.
2. Sallie Keller, US Census Bureau's Chief Scientist and Associate Director for Research &
   Methodology; Distinguished Professor of Biocomplexity, Biocomplexity Institute, University of
   Virginia.
3. Robin Bachman, Senior Advisor, Communications Directorate, US Census Bureau.
4. Karla Riso, Program Specialist, Deputy Directors Office, US Census Bureau.
5. Michael Thieme, Senior Advisor to the Deputy Director for IT and Operations, US Census Bureau.
6. Tamara Adams, Senior Advisor for IT and Operations. MOJO Lead, Office of the Deputy Director,
   US Census Bureau.
7. Deirdre Bishop, Chief, Geography Division, US Census Bureau.
8. Michael Ratcliffe, Senior Geographer, Geography Division, US Census Bureau.
9. Erika Becker-Medina, Chief, Decennial Communications Coordination Office, US Census Bureau.
10. Nate Ramsey, Director, Federal Statistical Research Data Center program, Center for Enterprise
    Dissemination (CED), US Census Bureau.

Meetings and Discussions with US Census Bureau Leaders and Experts.
Our collaboration with the Census Bureau provided important insights and knowledge. We appreciate
their time and care in sharing information about Census Bureau programs. The Curated Data Enterprise
language and ideas are beginning to be used at the Census Bureau and incorporated into the IT

Enterprise modernization. For example, Director Robert Santos (2023) has incorporated the CDE language into public statements, noting:

> We are moving towards a single enterprise, data-centric operation that enables us to funnel data from many sources in a single data lake using common collection and ingestion platforms. This enterprise-level data approach allows us to more effectively conduct our censuses and surveys. We will more selectively solicit data (e.g., survey data collection) to add value and accuracy. And when we solicit data from households, people, governments or businesses, we will be able to do so in a tailored, culturally relevant fashion, especially from those who have historically been the most challenging to secure participation. This is the **essence of a curated data approach — assemble, assess and fill in the gaps to create quality statistical data products.**

Through discussions with Census Bureau leaders, we learned about ongoing internal research and development of the linked universal frames, the Bureau's new centralized enterprise system, external data acquisition, the Annual Integrated Economic Survey (AIES), the FSRDC system, and many other topics. We participated in monthly and quarterly meetings. Census Bureau leaders also participated in our meetings and Listening Sessions.

A joint **Tabletop meeting** was held at the University of Virginia in Rosslyn to identify, discuss, and define the challenges for instilling the Curated Data Enterprise vision and concepts and the Census Bureau's Statistical Products First approach into the Census Bureau's IT enterprise system. Six high-level challenges were identified and the details for each were discussed. The consensus was that the most important was aligning new human processes with the linked business ecosystem.

Exhibit A1-1. Selected Meetings with US Census Bureau Leaders.	Dates
Christa Jones, Chief of Staff	Monthly
Robin Bachman, Senior Advisor, Communications Directorate	Monthly
Statistical Products First team (Erika Becker-Medina, Nathan Ramsey, and Emily Molfino)	Monthly
UVA and Census Bureau Business Ecosystem meetings led by Michael Thieme (over 20 Census Bureau leaders attend the meeting)	Quarterly
Deirdre Bishop, Director, Geography Division	Periodically
Jennifer Ortman, Demographer and lead on Demographic Frame	01/27/23
Erika Becker-Medina (partnerships)	02/14/2023, 04/7/23
Donna Daily, Division Chief, American Community Survey Office	02/15/23
Tom Mule, Continuous Measurement	03/14/23
Kevin Deardorff, Division Chief, Economic Reimbursable Surveys	03/24/23
Stephanie Studds, Associate Director for Economic Programs (3rd party data expert)	04/6/23
John Voorheis & housing team, principal economist, Center for Economic Studies	04/17/23
Omari Wooden, Census Business Builder	04/19/23
Jessica Wellwood, Business Frame presentation	05/9/23
David Beede and Nathan Ramsey, FSRDC	05/15/23
Jennifer Ortman, Demographic Frame presentation	05/23/23
Matthew Graham, Andrew Foote, Jobs Frames	08/29/23
Tammy Adams, US Census Bureau-UVA Tabletop Exercise – CDE Integration into Linked Business Ecosystem (20 Census Bureau leaders)	09/11/23

Exhibit A1-1. Selected Meetings with US Census Bureau Leaders.	Dates
Michael Ratcliffe and team, Geospatial Frame presentation	09/12/23
Jenna Morse, Blynda Metcalf, Edward Watkins, James Burton, Annual Integrated Economic Surveys	09/26/23
Lucia Foster and Randy Becker, Center for Economic Studies	09/29/23
Nathan Goldschlag, Center for Economic Studies	11/27/23

Appendix 2: Listening Sessions Summaries.

Appendix 2 provides a summary of each Listening Session.

The first two Listening Sessions addressed technical issues on developing use cases and Curation. The format for reporting is a summary of the discussion.

For the remaining Listening Sessions, the format mirrors the summary discussion in Part B of this report for each Listening Session individually. The following topics are presented: Purposes and Uses, Data Wish Lists, Infographics, Eliciting Purposes and Uses, and Collaborating with the US Census Bureau. These Listening Sessions were held with advocacy groups or those conducting resiliency research or sometimes both, such as the Listening Sessions with the National Low Income Housing Consortium (NLIHC) Disaster Housing Recovery Research Consortium (DHRRC).

Use Case Research Listening Session, June 9, 2022.

Background.
Use cases are a mechanism to engage the research community and promote methodological research to develop and implement the Curated Data Enterprise (CDE). Use cases involve discovering and integrating administrative records and third-party data with survey data to study key issues for the nation's communities. These steps require research on the viability of such records, record linkage technologies, and measures of uncertainty, leading to more accurate statistics about America's people, places, and economy. This listening session discussed ideas for developing a Use Case research program focused on use cases that can highlight and exercise the capabilities needed for a successful deployment of the CDE.

Six experts from local and state governments, universities, and Census Bureau economists discussed the criteria for and challenges of creating use cases to identify CDE capabilities. They were asked the following questions:
   o How to develop use cases as a mechanism to research the best ways to operationalize the CDE framework?
   o Who to engage the research community to identify and address research challenges?
   o What is the process to exercise and demonstrate the capabilities needed to realize the CDE concept to create a research program?

Discussion.
Several key themes emerged, particularly in the context of two Art-of-the Possible use cases[11] completed to date; these included the gig economy and migration due to climate-related events. To create viable Uses Cases, it was suggested that collaboration among various entities, such as national and regional Federal Reserve Banks, would be useful for creating a comprehensive view of economic activity and demographic shifts. This holistic perspective is essential for developing effective policies and support mechanisms.

Migration is another critical topic, often characterized as a "rare event" that poses unique challenges in data confidentiality and geographic understanding. The American Community Survey (ACS) plays a pivotal role in capturing nuances in small geographies, yet ensuring accuracy and confidentiality remains a complex task. Combined with other data, such as moving company data, cell phone records, and other administrative data, there is the opportunity to develop a demonstration Use Case.

---

[11] Art-of-the-possible use cases are focused on current policy issues for which there is no single or definitive data source to answer policy questions.

Climate-related initiatives also present opportunities for innovative use cases. For instance, understanding how businesses are exposed to risks such as sea-level rise or flooding can drive strategic climate responses. The Census Bureau is currently leveraging change of address information to better assess these risks, emphasizing the need for precise data on environmental exposure.

Developing use cases is challenging. One hurdle includes partnering with for-profit businesses regarding data access and determining the fitness of the data for specific uses. Moreover, as different datasets are integrated, setting priorities becomes crucial. Researchers must navigate issues related to noise and bias, particularly concerning traditionally undercounted or underrepresented populations.

One potential solution to these challenges is a meta-analysis of available datasets to ensure comprehensive coverage. This approach would facilitate a deeper understanding of the demographics and needs of various communities, including tribal populations, who may require targeted engagement strategies. Small area estimation is likely needed as well to ensure confidentiality.

Ultimately, successful use cases must encompass multiple Census Bureau frames (Geospatial, Business, Job, and Demographic), targeting at least three while ideally incorporating all available data sources. By addressing these considerations, researchers can create robust use cases that not only improve data accuracy but also enhance the understanding of the complex interplay between demographics, the economy, and other factors, such as the environment.

Curation Listening Session, June 22, 2022.

Background.
Ten curation experts from many fields participated in the Listening Session. Their fields included economics, sociology, behavioral science, biostatistics, statistics, digital archeology, and private sector companies, such as Deloitte, and data and modeling companies. They addressed the following questions:
- What is known about the curation topic? What are some key research challenges relative to the evolution of the CDE that are not currently being addressed?
- Who are relevant collaborators for executing a near-term research program to address challenges?
- Which researchers might be missed when thinking about curation for the CDE?
- Which aspects of the research is the Census Bureau well positioned to execute internally? Which aspects of research should the Census Bureau engage external partners?

Discussion.
The participants focused a lot on data curation, although we also nudged them to talk about curating the research cycle steps, i.e., the CDE framework.

Effective data curation is essential for maximizing the value of statistical information, particularly within the context of national statistics offices. To enhance curation efforts, we should actively tap into and collaborate with various international initiatives focused on open-source tools. Projects like SDMX[12] and SISCC[13] offer valuable frameworks that can be adapted to our needs, facilitating improved data handling and sharing.

---

[12] Statistical Data and Metadata Exchange – SDMX: **https://sdmx.org/**
United Nations description:
https://unstats.un.org/unsd/methods/itadvgroup/sdmx.htm#:~:text=SDMX%20%2D%20Statistical%20Data%20and%20Metadata,Nations%20Statistics%20Division%20(UNSD)%20and
[13] SISCC (Statistical Information System Collaboration Community) is a collaborative organization that provides online training and resources for capacity building in the Stat Suite and Data modeling in SDMX for data toolers and data producers. https://academy.siscc.org/

A critical aspect of this process is the investigation of platforms that manage accessible metadata for restricted data sources. Tools such as CKAN, DKAN, and DataVerse[14] can provide insights into best practices for metadata management, enabling us to better document and access data sets while ensuring compliance with privacy standards.

To streamline and standardize curation processes, several key R&D challenges must be addressed. Developing strategies that enhance the efficiency, transparency, and repeatability of key operations—such as acquiring federated and external data sources—will significantly contribute to overall effectiveness. Establishing robust metadata documentation standards is vital for managing the complexities of dynamic data sources. This includes creating fit-for-purpose standards that accurately reflect the processes affecting data quality, such as those involved in data collection, statistical processing, and disclosure avoidance.

Collaboration is also crucial for success. Engaging with diverse user perspectives, particularly from advocacy groups and marginalized communities with limited resources, ensures that our curation practices are inclusive and accessible. Partnering with state and local governments will further enhance our efforts, fostering a collaborative environment that supports the effective use of national statistics.

Curation of data, especially from historically underserved communities, requires building trust and ensuring that these communities feel their contributions are valued. Gaining access to their data necessitates a sensitive approach that prioritizes informed consent. Additionally, the role of various agencies in data curation must be transparent and well-documented, including standardization of metadata, particularly when integrating private data sources.

Effective curation must consider scalability; processes that work for one dataset may not be feasible for larger applications. It is essential to assess each data source's risks, errors, and trade-offs. Protecting the privacy and well-being of data subjects is paramount, as their identities and experiences can directly affect data quality and suitability for specific uses. This entails considering how sensitive data is handled and understanding that compatibility between datasets may be limited. Moreover, determining whether a dataset genuinely fits its intended purpose is crucial.

Navigating the complexities of data discovery can be challenging, particularly for those outside specialized fields, due to regulations on data sharing and confidentiality. Political influences also impact the availability and integrity of federal statistics, making curation a multifaceted endeavor that requires careful attention to ethical, practical, and regulatory considerations.

In summary, by leveraging international collaborations, investigating relevant platforms, addressing key challenges, and fostering inclusive partnerships, we can significantly improve our approach to curation, ultimately leading to higher quality and more accessible statistical information.

---

[14] CKAN (Comprehensive Knowledge Archive Network) is an open-source data management system developed by the Open Knowledge Foundation. It's used to power data hubs and portals, making publishing, sharing, and using data easy. CKAN has many features that other software like DataVerse may not provide. It also has a large community of developers and users who contribute extensions specific to CKAN. https://ckan.org/

DKAN is a Drupal-based open data portal built on CKAN. It shares many of CKAN's features, such as streamlining, publishing, and sharing data, but it is specifically designed to work within the Drupal ecosystem. DKAN is generally considered easier to configure and install compared to CKAN. https://dkan.readthedocs.io/en/7.x-1.x/introduction/index.html

DataVerse is an open-source data repository software, similar to CKAN and DKAN. It is widely used for open data. While DataVerse provides similar functions, it may not offer some features found in CKAN and DKAN. https://dataverse.org/

Local Data Partnerships Listening Session, December 6, 2022.

Background
The Curated Data Enterprise (CDE) aims to link data across different surveys over time, which was highlighted as a promising direction for creating more comprehensive and nuanced statistical products. The participants emphasized the need for more current, local data to inform community-level decision-making. These data include business and demographic characteristics, housing affordability, and public health metrics. We invited local and state data users to participate.

Discussion
Purposes and Uses.
Migration Data: What is real-time or near-real-time data on migration patterns these data are crucial for understanding the effect of events such as the pandemic, housing markets, workforce shifts, and regional economic planning.

Demographic Details: What are the characteristics of family units, including children's counts and ethnic diversity? Statistical products to inform social policies, educational planning, and understanding undercounts in decennial censuses and surveys. Such data can help in addressing disparities and ensuring a fair distribution of resources.

Gig Economy: What is a gold-standard measure of the Gig Economy? Detailed and timely economic data, including insights into the gig economy, are sought after for their relevance to policymaking, economic forecasting, and understanding labor market dynamics.

Supply Chain Analysis: What are supply chain dynamics at the local level? A consistent theme was the need for insights into supply chain dynamics at a local level. This highlights a broader interest in understanding how goods and services flow through economies and the impact of disruptions.

Access to Finance and Capital: How do small businesses access capital? The pandemic underscored the importance of understanding how businesses, particularly small and local ones, navigate financial challenges and access capital. This is crucial for economic resilience and recovery planning.

Workforce Dynamics and Migration: How do workers move through industries and across geographies? There was significant interest in the Longitudinal Employer-Household Dynamics (LEHD) program and its potential to shed light on workforce demographics, employer characteristics, and how workers move through industries and geographies. Understanding these dynamics is key for local economic development and planning.

Pre-Workforce Status Analysis: What is the status of individuals before they enter the workforce, e.g., school, military, etc.? This can inform policies around education, training, and veteran reintegration.

Tools and Infographics.
Interactive Geospatial Tools and Visualizations: These tools are sought for their potential to democratize data understanding and enable users to explore data in more depth, regardless of their level of expertise. This could include mapping socioeconomic trends, natural disaster impacts, or access to services, providing valuable insights for planning and development.

Data on Public Services and Infrastructure: A tool that addresses this question can inform strategies to bridge digital divides and promote inclusive growth.

Predictive Analytics and Scenario Modeling: There is a demand for data products that support predictive analysis and scenario planning, enabling users to explore "if, then" scenarios. This is particularly relevant for preparing for future events, such as climate-related disasters, and planning resource allocation or emergency responses.

Infographics and Customizable Outputs: Infographics that capture the imagination and democratize understanding, along with outputs that allow users to select variables and tailor data to their specific needs, were highlighted. Such products can make data more accessible to a broader audience, including those with less technical expertise.

Data Sources Mentioned.

Administrative Records and Third-Party Data: Integrating government and third-party data sources with traditional Census Bureau surveys is vital for enriching the statistical data ecosystem. This approach can enhance data products' granularity, timeliness, and relevance. While there's interest in integrating third-party data, including satellite imagery, there's also a call for transparency and caution about relying on proprietary tools. This highlights the need for a balanced approach that maximizes data utility while ensuring data integrity and transparency.

Survey Data: Traditional survey data remains a cornerstone, though there's a push toward making these data collections more efficient and responsive to emerging trends and needs. Pulse Surveys: Rapid response surveys for quick turnaround data were mentioned as a valuable tool, especially in capturing the dynamics of sudden societal and economic shifts, as seen during the pandemic.

Synthetic Data: Synthetic data, as illustrated in the nursing facility use case, shows how novel data approaches can fill gaps in understanding specific issues, such as staff's ability to reach work during emergencies.

Other Notes

Efficient Access to Data: Participants voiced a need for streamlined access to data, which would reduce the time and resources spent acquiring and processing data sets. This would enable more focus on analysis and the dissemination of insights.

Long-Term Vision for Short-Term Responses: There was a call for establishing long-term partnerships and frameworks to enable more agile and timely responses to emerging issues, such as emergency responses or understanding the evolving nature of work.

Third-Party Data Partnerships Listening Session, January 17, 2023.

Background.
Participants included chief economists and financial leaders in primarily large companies and the Atlanta Federal Reserve Board. The focus was on defining Purposes and Uses, and incentives to partner with the Census Bureau. The participants were excited about the possibility of combining data from the private sector with Census Bureau and other data, e.g., the Federal Reserve Board data, to generate granular and frequent Statistical Products.

Discussion.
Purposes and Uses.

Hybrid/Remote Work Patterns: What are hybrid and remote work patterns? The goal is to inform workplace policies and economic planning.

Personal Income Measurements: How can the accuracy of measuring personal income across various demographics and geographies be improved? The focus should be on improving policy targeting and economic analysis.

Population Movements and Patterns: What is real-time assessment of population movements for emergency preparedness, urban planning, and understanding migration trends? Data reflecting population movements, especially in response to climate events or economic factors, is critical. Such information aids in planning for emergency responses, urban planning, and understanding demographic shifts.

High-Frequency, Real-Time Data: What are near real-time insights into economic activities, consumer behavior, and societal changes? Such data are crucial for responding to fast-evolving situations like the pandemic, understanding spending patterns, and addressing policy issues.

Consumer Financial Health: What are insights into household financial health, such as checking account balances or asset movement, to inform policies on economic stimulus?

Incentives to partner with the US Census Bureau.
There's an interest in having data vetted or utilized by the Census Bureau as a mark of quality. This endorsement could significantly enhance the credibility and trust in data provided by third parties, making it a valuable benchmark for analytics and decision-making in various sectors.

- Ability to benchmark.
- Ability to track migration through loan applications.
- Ability to create a synthetic dataset based on consumer information (Census Bureau) and credit card records of purchases (private data) to understand consumer demands better.
- Monetizing data results.
- Federal agencies can bring together expertise in areas that private industry may also be interested in (i.e., supply chains).
- Consortium database for statistical purposes.
- User understanding of curated and authoritative data.

Participants highlighted the broader societal benefit of sharing data with government agencies like the Census Bureau. Contributing to national datasets can contribute to the public good, offering insights that aid policymaking, emergency response, and understanding societal trends.

Being recognized as a federal data analysis source can be a powerful branding advantage. It signifies the data's reliability and can lead to new business opportunities by demonstrating its utility in novel contexts or applications not initially envisioned by the data provider.

While there's acknowledgment that data sharing often involves costs, potential financial models could be explored to make the process sustainable. This might include direct compensation, shared revenue models from derived products, or other financial incentives that recognize the value contributed by third-party data providers.

The dialogue underscored the potential for mutually beneficial collaborations where statistical agencies and private data providers could pool expertise and data resources. This collaboration could lead to developing innovative data products, improved methodologies, and new insights into complex issues like supply chains, migration patterns, and economic trends.

The idea of **creating a centralized data repository or "data lake" was proposed**, where data from various sources could be pooled and made accessible for analysis. Such a model would require careful consideration of privacy, security, and access controls but could significantly enhance the ability to analyze and understand cross-sectoral trends and dynamics.

Use of State Early Childhood Integrated Data Systems (ECIDS) to address the Undercount of Young Children Listening Session, September 15, 2023.

Background.
We held this Listening Session with key state representatives who develop and use ECIDS-related data. Five states on our high readiness list (Salvo, Lancaster, Shipp 2024) participated: North Carolina, Minnesota, Iowa, Kentucky, and Iowa. Each state provided information on the data they curate and utilize. Early indications are that these states possess records from a variety of programs that may be helpful as part of the Census Bureau's demographic frame. Data from education, human services, birth, and related health records may improve coverage. For example, Minnesota's Early Childhood Longitudinal Data System (ECLDS) integrates data from education departments, human services, and health to provide a comprehensive view of early childhood data. It supports informed decision-making and policy development for early childhood programs.

High-quality data on children most likely to be missed in the decennial census and surveys, such as those who are poor or in need of government intervention, is important in correcting the undercount of young children.

It is impossible, however, to know how helpful such data may be until a sample of their records can be obtained for a proper evaluation. Participants identified a variety of administrative and legal hurdles that need to be overcome for a successful collaboration to occur. The first step is to develop data-sharing agreements with a university for aggregate statistics that can be compared with the 2020 Census. Participants from the states indicated that it would be easier to share their data with a university initially than directly with the Census Bureau, given the Bureau's more formidable legal and administrative requirements. Most importantly, despite the difficulties with the data-sharing agreement process, the state participants were excited to find ways to contribute their data as a potential pathway to improving the counts of young children.

Discussion.
Purposes and Uses.
Can State Early Childhood Integrated Data Systems (ECIDS) improve the accuracy and timeliness of counting young children? How do Public Health estimates of children in poor health change with more accurate data? Participants expressed a desire for statistical products that can represent young children in a timely manner, acknowledging the delay in when children appear in certain data sources and the need for more immediate data capture to better inform programs and policies.

Can the quality of the count of young children be enhanced using SNAP and WIC data? How does the acquisition and distribution of funds for childcare change with a more accurate count of young children? This approach would directly impact resource allocation, planning, and program effectiveness for young children.

Tools.

Head Start Performance Standards Data Set: This is a model for data collection and sharing.[15]

Direct Data Sharing and Analysis by Local Entities and States: States' potential to directly bring Census Bureau data into their systems, analyze it, and share results back was mentioned as a less legally complex model for collaboration, suggesting a flexible approach to integrating local insights into federal demographic analyses.

Data Sources Mentioned.

Early Childhood Integrated Data Systems (ECIDS) and local data:
States like Iowa, North Carolina, Kentucky, and Minnesota have systems in place that could potentially offer valuable, detailed data for enhancing the counts of young children. The discussion acknowledged the value and challenges of integrating this local-level data with federal efforts.

Minnesota's Departments of Education, Human Services, and Health:
For the ECLDS, showing a model of cross-departmental data integration for comprehensive early childhood data.

North Carolina's Early Intervention and CDE Data:
Discussed the complexities of sharing data with external parties, including legal and administrative challenges. An upcoming LDS (Longitudinal Data System) was mentioned, which could bring various data sources together.

Head Start Grantee Agreements in Iowa:
Highlighted as a significant achievement in data sharing, suggesting a potential model for collaboration and data integration. (See https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring)

Early Intervention and Pre-K Program Data: Discussed the complexities of tracking children's addresses over time, highlighting the potential for multiple addresses due to program participation at different ages and the need for point-in-time data accuracy.

Birth Records and Vital Statistics: Kentucky mentioned obtaining birth records, including the mother's mailing address, which could serve as a baseline for tracking young children and offer more complete data from the onset.

---

[15] See https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring.

Paid Parental Leave Program Data (Minnesota's Future Initiative): Mentioned as a potential future data source that could ideally capture children right when they are born, offering timely and accurate data for young children.

SNAP and WIC Information: Potential sources to improve the count of preschool kids, indicating a need to leverage administrative data for better demographic insights.

Administrative Data Sharing with the Census Bureau: Susan Brower, the Minnesota State Demographer, talked about the collaboration between state agencies and the Census Bureau for sharing administrative data, highlighting the value and potential of administrative records in enhancing Census Bureau data and other statistical products.

National Healthy Start Association Listening Session – December 15, 2023.

Background.
NHSA's Mission is to be the nation's voice in providing leadership and advocacy for health, services, and interventions that improve birth outcomes and family well-being. (https://www.nationalhealthystart.org/about-us/) With over 100 Healthy Start projects across the United States, the Association works to expand the capacity of community-driven and community-based MCH programs and maternal and infant mortality preventive health services, thereby ensuring that all families have access to a continuum of affordable quality health care and related services. This range of care extends beyond the welfare of newborns and includes mothers, fathers, and families, thus affecting the entire community.

NHSA's leadership team, including its data scientist, attended the listening session (https://www.nationalhealthystart.org/our-team/). They identified the following purposes, uses, and tools that would facilitate their work.

Discussion.
Purposes and Uses.

   What are the community characteristics at the neighborhood level? Participants use the American Community Survey (ACS) to develop community profiles for their programs. "I love the ACS, but working with the ACS requires a lot of work. The numbers don't tell the full story. We still must add the context and connect the dots. What would be useful are data stories and visualizations to communicate the full story to the community and to write grant proposals."

   What is the relationship between the Social Determinants of Health and race on health outcomes? Participants found previous CDC data highlighting how race impacts health outcomes to be tremendously useful for connecting the dots between social determinants and health outcomes. This data helps in understanding the broader context of health disparities and in developing targeted interventions.

   What is the effectiveness of early intervention programs, e.g., Are children's developmental milestones on target? They expressed interest in the potential use of Early Childhood Integration Data Systems (ECIDS) to provide comprehensive insights into the health, education, and social service interactions of young children.

Tools.

> **Updated and Simplified Data Visualization Tools about Infants and Children:** Data visualization and obtaining data for specific populations within communities, such as the perinatal infant population in certain regions, need to be simplified. Simplified visuals and access would enable community organizations to better understand and utilize data for program planning and evaluation.

> **Data Templates for Community Organizations:** The idea of creating data templates that community organizations can use for specific purposes was discussed to help translate complex data sets into formats helpful to communities. This could assist in grant applications and program planning by making Census Bureau data more navigable for organizations without statistical experts.

> **Improved Counts of Young Children:** Participants are interested in more accurate and comprehensive counts of young children to enhance resource allocation, program planning, and evaluation. They recognize the limitations in current data collection methods and the impact of undercounts on service delivery and effectiveness. Accurate counts are essential for determining the scope of need and measuring the impact of interventions on target populations, particularly in addressing disparities and improving maternal and child health outcomes. "What keeps us up at night is worrying about how accurate the denominator is – the count of young children."

Data Sources.

> **CDC and WIC Data:** Supplemental data sources like WIC are used to compare and augment Healthy Start data, especially in understanding breastfeeding trends and other health indicators within similar demographics.

> **American Community Survey (ACS):** Used by participants to develop community profiles and understand social determinants of health, albeit with some challenges in terms of data complexity and the need for additional context.

> **Community-Based Data:** The potential for leveraging local data to supplement traditional Census Bureau sources was discussed, indicating a need for more flexible data collection policies and improved infrastructure at the community level to capture more accurate counts to understand the needs of underserved populations.

Other Notes.

> Participants expressed interest in engaging more directly with Census Bureau efforts, specifically mentioning the potential for collaboration with the Young Children Working Group. The idea of having a dedicated liaison or point of contact within the Census Bureau to assist with navigating data resources and disseminating information was highlighted as beneficial.

> The need for technical support, possibly through an online chat option for real-time help, was mentioned as a practical way to support community-based organizations in utilizing data more effectively. *(Note, this could be one part of a proposed Data Concierge service discussed in the main report.)*

> The session concluded with suggestions for the Census Bureau to re-imagine its approach to data dissemination, particularly in providing tailored support and resources to organizations with varying levels of data literacy and capacity.

Community Health Advocates – December 21, 2023.
Robin Bachman, Senior Advisor, Communications Directorate, US Census Bureau, introduced us to Will Snyder, co-founder and CEO of Metopio. He brought together public health officials and analysts who use his data system. Metopio's vision is to make ACS and other data sources easy to access and use and create tables and visuals to prepare Community Health Assessments and other community-based reports to assess a city's or county's health. Metopio has built a data system that uses ACS, CDC, IRS, EPA, and other data sources. They survey about 25% of hospitals in the country.

Metopio's mission is to make it easy for public health officials and analysts to access data for creating Community Health Assessments, also known as Community Health Needs Assessments, every five years. A CHA provides a comprehensive picture of a community's current health status, factors contributing to higher health risks or poorer health outcomes, and community resources available to improve health.

Community health assessments comprise data and information from multiple sources, describing the community's demographics, health status, morbidity and mortality, socioeconomic characteristics, quality of life, community resources, behavioral factors, the environment (including the built environment), and other social and structural determinants of health status. See Public Health Accreditation Board, *Standards-Measures-Initial-Accreditation-Version-2022.pdf (phaboard.org)*, October 2022.

The Metopio data are also used to assess community health in interim periods. Their mission statement aligns with the Census Bureau's Statistical Products First approach.

- We've set out to help anyone and everyone feel like a data expert.
- Simply put, better decisions can be made with better data. That's why we make tools that make it simple to curate, analyze, and visualize data.
- With Metopio, you can find data on any place and population, put it into context, and visualize the findings in ways that almost anyone can understand.

Metopio has added an AI tool to facilitate searches. We introduced Emily Molfino, US Census Bureau, to the CEO, Will Snyder, and they have met to discuss this feature.

Metopio compiles these estimates for many public health districts. However, data gaps remain, as there is a need for data at lower geographic levels, below the tract level, for specific groups, such as African American households aged 65 and older in a 6 x 6 block area.

Discussion.

Purposes and Uses.

What are the characteristics of high-need communities? What challenges do they face accessing services? The Listening Session participants focused on **high need communities**. These communities require support for food, transportation, healthcare, and other services. Integration of services can more efficiently and help these groups to thrive. Metopio defines these communities using a Social Vulnerability Index, a hardship index, and other factors, including non-medical social needs that can affect health. Other data sources include EPA environmental data, data on life expectancy rates that have been declining for the last ten years, and other emerging issues, such as the impact of climate change on health.

Where are service deserts? What distances and challenges do children and older people face in obtaining health and related services? They highlighted the need for geospatial data

products at the block level to help identify service deserts and effectively plan resource allocation. These data would support efforts to ensure access to services for young children and vulnerable communities. Block-level data and distance-related tables would aid in more precise planning and implementation of services at the community level.

What are the social determinants of health for the uninsured population? What are their needs and challenges?

What are older adult African Americans' demographic characteristics and health status in a specific Census Bureau tract and block? What are the community assets and gaps in delivering health services to a community? Can tools be developed to integrate CDC health data and decennial and ACS demographic data to overcome the complexity of merging these data sets manually? Modeled or synthetic data would be acceptable to achieve granularity on CDC information, particularly for small area estimates. *Additional Stratifications*: To assess health disparities, health and demographic data of students by school enrollment by race and ethnicity for preschool, K through 8, and high school.

Data Sources Mentioned.

Data from Local Community Organizations: Data from daycare centers, pediatric clinics, and community health programs could provide alternative counts or insights into the population of young children.

Aggregate Health Record data: Utilized both for understanding health trends among young children and potentially to estimate population counts where Census Bureau data may be lacking.

School Enrollment Data: As a proxy for understanding the demographics and numbers of children in a specific area, this could help adjust for undercounts in Census Bureau data. (State Longitudinal Data Systems (SLDS)).

CDC Data: Utilized for health-related information, although noted as less granular compared to other sources.

ACS (American Community Survey) Data: A primary source for demographic data but identified as cumbersome for detailed population group analysis.

IRS Data: Mentioned for augmenting or supplementing household wealth characteristics that are missing from the Census Bureau data. Specifically, IRS data could enhance health insurance marketplace enrollment understanding, including the ACA marketplace credit and age stratifications.

Tax Return Data: "robust, little room for people to lie."

Language Assessment Data: Identified as critical for understanding diverse community needs, particularly for hospitals and health departments to meet regulatory requirements. However, current data is outdated and lacks granularity at desired levels (e.g., block and block group levels, ZIP and sub-ZIP levels).

Data Wish list.

Demographic data stratification needs to be more granular in terms of types and geography. For example, the Asian category is too broad, as there are differences when studying Vietnamese, Korean, etc. In many school districts, students speak 30 to 90 different languages.

Synthetic data – there is a need for the Census Bureau to fill that gap. Companies can only access public use files, so their models and estimations are not as detailed.

Identifying community assets to inform community priorities.

IRS Section 501r reporting[16]**:** hospital organization uses Form 990, Schedule H, Hospitals, to provide information on the activities and policies of, and community benefit provided by, its hospital facilities and other non-hospital health care facilities that it operated during the tax year. A specific requirement includes doing a language assessment of hospital patients, which may require modeling to estimate.

School enrollment data (available for K-12) are helpful but do not include race and ethnicity for preschoolers.

Age data by year rather than groups in the American Community Survey.

Tools.
- Over 5,000 hospitals and every public health department seek qualitative information from their residents through town halls, focus groups, and surveys to create a Community Health Assessment, collecting data in various ways. Could the Census Bureau use Machine Learning techniques to process and synthesize the qualitative data? Would the survey data be useful to the Census Bureau, and how?
- Guidelines on how to conduct local surveys.
- More tools to help work with ACS data which "are really quite messy" and have more changes year to year than expected. Specific guidelines would be helpful too, e.g., using the 5-year ACS that covers the COVID years versus the non-COVID years.
- Standardize social vulnerability indexes or create several indexes and explain the differences.
- Credit scores, debt history, and third-party commercial data to score or assign risk as another way to measure high need communities.

Eliciting Information about Purposes and Uses and Ensuring Broad Participation.
- Implement a feedback mechanism on data utilization platforms to understand user intentions and requirements.
- Working intentionally with community-based organizations with deep connections to the communities of interest ensures diverse voices are heard.
- Validating findings and data interpretations with the communities they represent to ensure accuracy and relevance.
- Establishing guidelines for surveys to ensure they capture the necessary information for the Census Bureau and other stakeholders in a manner that is inclusive and representative of all community segments.

Community Solutions – January 16, 2024.

Background.

---

[16] 501(r), also commonly called financial assistance or charity care, is a law requiring nonprofit hospitals to give patients discounts based on their household income.

The Center for Community Solutions is a nonprofit, nonpartisan think tank improving health, social, and economic conditions in Ohio through policy advocacy, research, and data analysis.[17] The Center is a trusted source of analysis and advocacy with a century-long legacy of pragmatic problem-solving in Ohio's health and human services landscape. The Center has three focus areas – policy advocacy, research, and data analyses. Their primary areas across these focus areas are maternal and infant health, Medicaid, poverty and the safety net, and older adults.

Robin Bachman, Senior Advisor, Communications Directorate, introduced us to Loren Anthes, a Principal at Health Management Associates. Loren then introduced us to Emily Muttillo, the Director of Research, Williamson Family Fellow for Applied Research, who worked with us to organize the session.[18]

A diverse group of participants from Urban Institute, Cleveland University, Legal Aid Society of Cleveland, Cleveland Foundation, Food Bank, Health Policy Institute of Ohio, Ohio University's Heritage College of Medicine, UVO Health, LEAP, Community Education, and Neighborhood Leadership Institute. The Census Bureau participants included Robin Bachman, Senior Advisor, Communications Directorate, and staff from the Office of Strategic Alliances (OSA).

Discussion
<u>Purposes and Uses</u>
Who is disabled according to the federal definition of disability?

Defining the population with disabilities more completely.
- There is a lack of disability data to examine the impact of disabilities on individuals. The first issue is how do you define if someone is disabled.
- We need data for policy and funding requests, but the disabled population is undercounted. The **ACS is good, but it it's limited. It leaves out a significant population of people with disabilities**, such as people who are incarcerated, people who live in institutions, people with neurological disabilities, people with psychiatric disabilities, people with chronic disease, and elderly people who say, "I am losing my eyesight." We can use Medicaid Medicare data. These data are current but not complete. Neither data source includes all disabilities. The **Federal definitions of disability are based on eligibility for services, not based on the need or functional limitation of the individual who could benefit from services**.
    - The federal definition of disability is a physical or mental impairment significantly limiting a person's ability to perform one or more **major life activities.** It can also include having a history or record of such an impairment or being perceived by others as having such an impairment, e.g., a scarred face that makes it difficult to find a job.
        - **Major life activities** are those functions that are important to most people's daily lives. They include breathing, walking, talking, hearing, seeing, sleeping, caring for oneself, performing manual tasks, and working. Major life activities include major bodily functions such as immune system functions, normal cell growth, digestive, bowel, bladder, neurological, brain, respiratory, circulatory, endocrine, and reproductive functions.
- People with disabilities often don't identify as having a disability. They say that they have a chronic health condition. About 40% of disabled individuals are not counted. This is a concern for advocates who are advancing policy and program changes.

---

[17] https://www.communitysolutions.com/
[18] https://www.communitysolutions.com/team-members/emily-muttillo

- Public health agencies do not collect disability data. "We want to understand the (1) need for food, housing, and transportation, (2) their chronic health and disabilities, (3) and access to health care."
- The ACS question on disability is one of the better ones because it asks about a series of components, and then you compile disability measures. And yet, the Census Bureau is moving away from that.
  - The disability community is not so concerned about expanding the questions. The question is how is the Census Bureau considering somebody to have a disability based on their responses? It's a significant concern because it will dramatically impact the number of people with disabilities who are counted. The disabled population is already undercounted. The ACS questions are a very good tool, but they undercount people with disabilities by half. There is concern about changing the question. We will lose continuity, but it might further undercount people with disabilities. There are concerns about the Washington Group's surveys (https://www.washingtongroup-disability.com/).
- The US Health and Human Services (HHS) promotes the ACS questions to be included in public health data. This will be the first time over the next decade or so that we can have comparable data about disability. Changing the ACS now, would be a real challenge.

Who has access to transit versus who uses public transit? What is the relationship between drivers' license suspensions and use of public transit?
- Provide insights into economic mobility and how commuting patterns affect local economies and individual livelihoods.
- Accessibility and Mobility Data: Use transit data or networks to better understand who has access to transit rather than just who is using it.
- Data on driver's license suspensions could provide insights into mobility, as a significant portion of the adult population with a suspended license may have a vehicle available but cannot legally drive. Depending on the state, between 5 to 17% of drivers have suspended licenses. These data sources could add value to understanding accessibility and mobility issues within communities, which is crucial for urban planning, public transportation policy, and social services provisioning.

What are food bank outcomes on reducing food insecure households?
- What are other sources of food that food-insecure households get?
- How do you measure the number of food insecure households?

What are local literacy rates, school enrollment, and employment trends?
- These data are needed for planning and development purposes to address educational needs and to understand employment trends for better workforce planning.
- Potential data source: State Longitudinal Data System.

What types of legal problems do low-income (poor) people have?
- Note: The Legal Services Corporation (LSC) is federally funded through congressional appropriations. The LSC is a non-profit corporation that provides civil legal aid to low-income Americans. The LSC receives most of its annual funding from the House and Senate Appropriations Subcommittee on Commerce, Justice, Science and Related Agencies.
  - They do not see all the problems because people do not know that legal aid is available.
- LSC periodically runs the **Justice Gap Survey** to measure the prevalence of various kinds of legal problems, especially for the poor.
  - When asked about legal problems, do respondents perceive that a problem that they experience in their lives is a legal problem?

- Are there other data sources that could be used to provide data more frequently and at more granular geographic levels? Some ideas discussed include eviction data, court data, or driver's license suspension (due to unpaid debts), which are available in many states. These data do not include demographic data so need to interpolate by Zip Code.
- Other challenges, such as losing Temporary Assistance for Needy Families (TANF) benefits, are harder to discern.
- This raises a broader question of finding consistently reliable data for other populations, e.g., LGBTQ.

Data Wish List.

The Office of the National Coordinator has released new interoperability standards for electronic medical records (EMRs). Can social conditions and risks be more comprehensively connected?
- The new standards released by the Office of the National Coordinator regarding interoperability standards for EMRs, specifically HL7 FHIR standards. The idea proposed involves creating an underlying architecture or resource library associated with Census Bureau data that could map to HL7 FHIR resources (https://www.hl7.org/fhir/overview.html). This would facilitate communication between medical resources used in medical settings and community-based organizations, contributing to a better understanding of social conditions and risks more comprehensively. The aim is to use HL7 FHIR standards to improve the integration of medical data with Census Bureau data, thereby enhancing the ability to understand and address social determinants of health effectively.
- For example, an individual has a medical appointment. The patient has conditions that have been identified. Could synthetic information (or H7 FHIR) transmit that type of information to the Census Bureau to get a more holistic understanding of the patient? It would be very useful and powerful to provide community-level data in a medical context since Medicare, Medicaid, and others are moving a lot of the money that way.

Could the ACS and CDC Behavioral Risk Factor Surveillance System (BRFSS) combine forces to make both data sets more useful? Is this an opportunity for Small Areas Estimation and privacy protection?

ACS questions about gender and sexuality are going in the right direction. The write-in options are a best practice for inclusivity if the multiple-choice answers do not fit the person.
- Why is the cut-off age set at 15? This feels arbitrary.
- Does the person filling out the ACS survey on behalf of their household know the sexual orientation or the gender identity of everyone in their household?
- Many people do not feel safe reporting on their sexual and gender identity. Building trust is important.
- The Trevor Project (https://www.thetrevorproject.org/) provides data on LGBTQ youth statistics. Pew Research Center provides good data on adults (https://www.pewresearch.org/topic/gender-lgbtq/).

Tools and infographics to educate data users.

Undercount of Young Children.
- Half of that problem is that people think the decennial census is only for adults. They just leave children off. They are not aware of who should be included.

- One way to fill the gap is to use Administration for Children and Families administrative data. This involves negotiating for the data with each state.
- State Early Childhood Integrated Data Systems (ECIDS) and State (K-12) Longitudinal Data Systems (SLDS) were also mentioned.

Custom Geographies.
- We have had to create our own geographies for Cleveland. These include neighborhoods and wards in the city. Community Solutions has developed a methodology in collaboration with Cleveland State University.
- There is a need to create custom geographies to focus in on a local population for service, provisions, funding, etc.

How do we elicit information about purposes and uses fairly?

- Building trust and focusing on strengths in communities.
- Working with the Neighborhood Leadership Institute, we learned that our data **presentations were "problem and deficit-focused instead of strength-based."** This approach is not good for trust-building. We need to focus on opportunities.
  - Small sample sizes make it challenging to work at the community level.
  - One issue is that synthetic data are sometimes difficult to explain to policymakers.
- When we work at the community level, we have data factsheets by ward and neighborhood to facilitate conversations with the community. The data are depressing and not inspiring. The challenge is explaining the data in a compelling way.
- Building websites to make definitions and data accessible is best practice. A good example is the Robert Woods Johnson Foundation website.
- It is critical to involve the community in decisions about changes to the survey. For example, people with disabilities were not included in the ACS changes being discussed.
- Is it possible to target or ask supplemental questions, e.g., Is gentrification an issue for your neighborhood?

National Low Income Housing Consortium (NHLIC) Disaster Housing Recovery Research Consortium Coalition (DHRRC) Listening Session – a mix of advocates and researchers – May 15, 2024.

Background.
NLIHC is dedicated to achieving public policy that ensures people with the lowest incomes have quality homes that are accessible and affordable in communities of their choice. Their focus areas are to educate lawmakers and the public about the need for affordable homes, to mobilize members and supporters across the country to advocate for good housing policy, and to shape public opinion about low-income housing issues. "The DHRRC is organized around key principles to ensure a just recovery for the lowest-income and most marginalized disaster survivors and their communities."[19]

We worked with Andrew Aurand, the senior vice president for research at NLIHC. He leads the DHRRC research team that documents the housing needs of low-income renters in annual publications like *The Gap* and *Out of Reach.* He also conducts additional research that informs housing policy related to extremely low-income renters and co-manages the National Housing Preservation Database. He coordinates quarterly and annual meetings with the NLIHC DHRRC, which includes advocacy organizations and researchers, to identify data and research needs across the groups.

---

[19] https://nlihc.org/disaster-housing-recovery-research-resilience

The group included about 30 people (including NLIHC staff), and our Listening Session was integrated into their annual meeting. We held a group discussion and then broke into four breakout groups facilitated by US Census Bureau staff, two for advocacy organization participants (Chip Walker and Karla Riso) and two for research participants (Bethany DeSalvo and Robin Bachman).

Discussion.
Purposes and Uses.

Resiliency: What are the characteristics of households most vulnerable to flooding by characteristics and location? What kind of insurance do they have? Who does not have insurance?
 Focus on impoverished renters since they bear the brunt of disasters. These data would be used to create short and long-term preparedness plans in flood areas.
 For homeowners with mortgage paid, what types of property insurance do they have and how much do they pay in property taxes?
 There is an insurance crisis in Florida, Texas, and California and across the country, see article (New York Times: "Insurers Around the US Bleed Cash From Climate Shocks, Homeowners Lose." https://www.nytimes.com/interactive/2024/05/13/climate/home-insurance-profit-us-states-weather.html.

Mobility and displacement – Can mobility and displacement be tracked using HUD assistance for individuals and households at the tract level (or lower)?
- Who resides in an area before and after a disaster? Where do people go? Who returns?
- Why do rents increase after a disaster? Is it rent gouging? Have costs increased, such as insurance, property taxes, etc.?
- Why are people moving to high-risk climate areas?
- How do FEMA rules affect Puerto Rico recovery from disasters? FEMA claims only apply to the main house on a large piece of land with many homes, e.g., a family enclave with many homes and uses of the land across family members.

How does redlining, the history of discrimination in appraisals, mortgages, etc., affect communities' resiliency today?

How do you define housing recovery after a weather or other event? How are people in a disaster area moving through the system? Are resources being deployed in a timely way? Are renters able to repair enough to make the place livable?

Where are low-income communities in distress? Low-income homes, households, and communities experience more damage and are less likely to see recovery. What blocks are prepared/not prepared for disasters? What are timelines? Short-run displacement? Longer run – what happens to households? *(FEMA talked about how many households are still in FEMA trailers at $50 per month three years after a hurricane.)*

Are communities designed to minimize heat? e.g., NOAA shows where concrete is. What are access and affordability issues related to heat, air conditioning, broadband, and utilities?

How to make homes more resilient? Are buildings and repairs up to code? Or should they be built better than code, so the structures are stronger to sustain climate events?
- Note: NOAA heat maps are used as a proxy for potential disaster-prone areas; see Census Bureau's My Community Explorer, Extreme Heat Events, https://disasters.census.gov/pages/heat-events.

Does Green infrastructure save money? What are the timelines?

Tools.

Provide tools to build administrative capacity for local communities, especially small and rural communities. Need this for pre-disaster, not just post-disaster.

Bring the American Factfinder back. The new data tools are much harder to use.

Methodologies and tools to work with the data, e.g., synthetic data.

Geographies – block level <10 people; Census Bureau tracts are often too big, e.g., bigger than a county.

Data Wish List.

More data for rural areas and unincorporated areas (no building codes or other regulations).
Discover, provide, and maintain uniform data coding to facilitate comparisons across states.
- Housing valuations.
- Building codes for each municipality.
- Permits by types.
- Roof-level data to show the condition of the home.
- Coastal and riverine flooding data.
- Loans and other funding to survive transitions from a climate event to finding stable housing, employment, etc.
- Potential data source: Army Corps of Engineers, National Structure Inventory (NSI).

References were provided to participants.
- Census Bureau Disaster data main page: https://disasters.census.gov/pages/other-events.
- Transformation of the Census Bureau Business Ecosystem. https://www.census.gov/about/what/transformation/maximizing-operational-efficiency/data-centric-business-ecosystem.Overview.html#list-tab-1020552672.

International Downtown Association (IDA), June 12, 2024.

Background.
IDA connects and empowers place management leaders with knowledge, research and public policies for creating prosperous city centers, commercial neighborhoods, and livable urban places for all.[20] IDA conducts many activities for its members, including leadership development, organizational management, economic development, policy and advocacy, and more. Their *Value of US Downtowns and Center Cities* report shows a downtown's impact within the context of its city and region through the lenses of Economy, Inclusion, Vibrancy, Identity, and Resilience.

Their Vitality Index is an interactive, online tool for benchmarking the vitality of downtowns across the United States. The Vitality Index builds on IDA's pioneering research in *The Value of US Downtowns and City Centers to measure vitality through three principles: Economy, Inclusion,* and Vibrancy. Fifteen people from a variety of cities and IDA staff participated in the listening session. The cities included Memphis, Nashville, Baton Rouge, Detroit, Philadelphia, Charlotte, and Houston, and two companies that provide analytical services to the cities – VibeMap and MJB Consulting.

---

[20] https://downtown.org/about-ida/

Discussion.

Purposes and Uses.

How many people work in downtown areas? Is it possible to differentiate visitors from workers?

Current Process:

- Use cell phone intelligence data, e.g., AI-Placer data.
- Use to rely on LEHD. Time lags for releasing the data make it less useful.
- Current data do not count significant federal presence, e.g., the Department of Homeland Security, the FBI, or the Department of Justice.

What is the distribution of current and emerging talent? Talent acquisition – who is there, who is coming, and what is the difference between aspirational jobs and attainable jobs? What is the path from attainable to the aspirational job? Are certifications, community college degrees, and other training credentials indicators of current and emerging talent supply and demand? (NLx data).

What are the characteristics of Activity Districts in a downtown area? For example, what does a downtown need to know if it is building a stadium? How does this change the economic and social characteristics of an area? Does it attract businesses, new residences, or other spillovers (positive and negative)? What is the radius of the impact (1/4 mile, ½ mile, or more)?

What are the characteristics of resilient downtown areas?

- Multi-use buildings (commercial, retail, hotels, residences). Note that residential growth is booming, which will maintain downtown resiliency.
- Are there sufficient schools, daycare, parks, and sidewalks to attract families.?
- Environmental resiliency – Connect data sources: walkability, reduced car usage, increased public transportation use; bike lanes.
- Community engagement – workers and residents – use of retail, attendance at sports and concerts, etc.
- Urban density – is it inclusive? Does it include low-income households?
- Review State of Downtown Reports (examples can be provided).

Who is moving in to downtown areas? Who is moving out? Where are they coming from/going to? Are families moving to downtown areas? What are household compositions and ages? Placer.ai (https://www.placer.ai/home-new) cell phone data, and Arrivalist, (https://www.arrivalist.com/) were mentioned as possible data sources.

How many people are Unhoused? What are trends (changes) over time? What types of housing are available for the unhoused? What services are available for the unhoused?

What kinds of retail should a downtown area develop? What are population characteristics within a 5-mile radius of downtown, e.g., percent with a car, have three or more people in their household, and have high incomes? What makes the area attractive to develop?

Do people stay in their neighborhood or go outside their neighborhood to shop (called sales leakage)? Compare spending by category, such as restaurant revenue, to actual gross revenues. Are dollars leaking to other areas?

What are clusters of downtowns that are similar? For example, what are five other downtown areas that are most like St. Louis? Create an index to facilitate comparisons. For example, the two downtowns are not similar if St. Louis is 100 and Miami is 45.

- The analysis could initially be done at the MSA level and downtown areas.
- IDA's Value of Downtowns is a useful tool to build on. Index values of inclusion, vibrancy, resilience, and typology (based on the citywide share of average growth in employment, density, population, and assessed value groups downtowns as established, growing, or emerging).

Tools.

ACS trends. There were requests for the Census Bureau to provide these data. It is time-consuming for researchers and data users to do this work.
- Data Users must first download ACS tables each year, strip unwanted information, and prepare calculations to show trends over time at the Census Bureau tract level.
- Time-consuming for routine tasks (many hours).
- Challenging to match Census Bureau tracts over time.

Geographies:
- **Custom geographies to create data for the downtown area at the block level for ACS, County Business Patterns, and other surveys.**
- Create trend data with the challenges worked out, e.g., matching Census Bureau geographies over time.

Use non-Title-13 data to create statistical products with non-Title-13 data, e.g., a map of all businesses, housing units, and other characteristics of interest.
- o The National Establishment Time Series (NETS) Database currently contains a database of current businesses by year, Black Knight data, and other third-party data sources.

Create a tool to show data for all types of work arrangements by downtown areas, e.g., commuting by car, transit, bicycle, working from home, or hybrid by the number of days working in an office.

References

City-level research at Brookings (William Frey), June 6, 2024.
https://www.brookings.edu/articles/census-shows-big-city-population-improvements-countering-predictions-of-the-urban-doom-loop/.

Downtowns Rebound: The Data-Driven Path to Recover, https://centercityphila.org/research-reports/downtowns-rebound-report-2023.

Center for Risk-Based Community Resilience Planning: A NIST Center of Excellence, led by Colorado State University, Listening Session, June 18, 2024.

Background.[21]
The Center Team comprises more than *90 individuals*, including researchers, programmers/developers, NIST collaborators, postdoctoral scholars, and graduate students from our University Partners. Working closely in teams on more than *40 tasks*, the Center of Excellence will provide a common data architecture by collaborating with the National Center for Supercomputing Applications to ensure that data from around the world can be seamlessly integrated into a robust computational environment known as IN-CORE. IN-CORE will allow users to optimize community disaster resilience planning and post-disaster recovery strategies intelligently using physics-based models of interdependent physical systems combined with socio-economic systems.

The ability to comprehensively model community disaster resilience requires that experts from several disciplines work in concert to systematically model how physical, economic, and social infrastructure systems within a real community interact and affect recovery efforts. Currently, no models consider all aspects of how a natural disaster affects a community or measures its resilience quantitatively.

The Center for Risk-Based Community Resilience Planning uniquely merges the disciplines of Engineering, Social Sciences, and Economics to comprehensively model community resilience. Systems that are essential for the recovery and vitality of a community, such as technological, financial, social, and political support, healthcare delivery, education, and public administration, are being integrated in the model, creating a nexus between social and technological infrastructure networks that will narrow the gap between engineering and social science aspects of resilience planning and will facilitate risk communication among stakeholders and community resilience planners. The work products from the Center will provide a quantitative and science-based approach to community resilience assessment and, for the first time, will support a business case for enhancing disaster resilience at the community level.

Full validation of the system architecture in IN-CORE will be possible through extensive field studies focused on community resilience and recovery rather than infrastructure damage and failure studies. IN-CORE will be able to answer detailed questions on the lingering effects of natural disasters on communities, population dislocation, health and the well-being of the residents, impacts across the economic spectrum, and fiscal impacts, thereby assessing community resilience and disaster recovery via a suite of resilience metrics.

Discussion.
Purposes and Uses.

What are the effects of climate events on population movements and employment? NOAA has data and models that more accurately measure migration. Other potential data sources include USPS Vacancy data. https://postalpro.usps.com/ot. Researchers would like access to these files to understand migration. However, it appears that only companies can access them. The IRS also accesses these data but only makes it available at the county level.

What are the impacts of disaster on multifamily housing and insurance coverage data?
Census Bureau data provide context about the housing market, e.g., ACS, AHS, and decennial. They do not provide tabulations about buildings and properties as units of analysis. There are some cross-tabulations for tabulating households by structure types but does not provide building information. The

---

[21] http://resilience.colostate.edu/index.shtml.

Rental Housing and Finance Survey, https://www.census.gov/programs-surveys/rhfs.html, provides some information but at a national level, not at smaller geographic areas. HUD and FEMA have datasets about disaster recovery funding but there is generally little to no data about property insurance coverage. ACS has a few questions about hazard insurance, but other types of insurance information are needed, e.g., the amount of insurance premiums and what is covered. They were able to obtain Texas wind insurance data. CoreLogic and Black Knight do not have insurance information. *This is an opportunity for data discovery to identify what third-party and government data sources are available*.

Are US Geological Service (USGS) natural resource data to identify useful for predicting flooding, wildfires, and other climate-related events?[22]
There is a wealth of detailed natural resources data with metadata, e.g., everything related to water. *The Census Bureau does work with them.* The information would be valuable for disaster planning and recovery. USGS holds public meetings and forums to invite input into how they do their work. See National Geospatial Act of 2018, https://www.fgdc.gov/gda/geospatial-data-act-of-2018-summary.pdf.

How should resilience be measured? There are dozens of definitions, and very few have been empirically tested. Why should one definition be used over another? Federal Reserve Banks (San Francisco) and National Labs have worked on these definitions (ANL, INL, ORNL).
- Argonne's National Economic Research & Resilience Center provides "innovative and transparent economic data access, research, and analysis to support and strengthen economic development and resilience in communities across the nation." It worked with FEMA to create the National Resilience Index.
- "Idaho National Laboratory's Resilience Optimization Center (INL ROC) tackles infrastructure resilience challenges through applying laboratory-wide capabilities and expertise." https://resilience.inl.gov/
- "ORNL is taking an integrated and proactive approach to including sustainability efforts and climate change resiliency into the operational planning process. The Climate Change Resiliency (CCR) Roadmap is an important part of the SCI Roadmaps Projects Structure. This roadmap includes the ORNL CCR team, which was chartered in FY 2014." https://www.frbsf.org/topics/climate-resilience/.

Resilience is of interest in terms of how populations recover from disasters. The San Francisco Federal Reserve Board also has an economic definition of resiliency. Some approaches include a resiliency focus on different aspects, e.g., household, community, and employment – especially low-income households with respect to these characteristics. Where do these households go? Low-income households often never recover after a disaster.

Infographics
Description of when to use one-year and five-year ACS data. What are the differences between the two? Why are the estimates different? What should be used for population projections? Provide *What-if* scenarios.

Tools.
Create a tool replicating the CDC Social Vulnerability Indexes. Allow users to add or delete variables.
CDC Social Vulnerability Indicator (SVI) is composed of 16 variables, https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

---

[22] https://www.usgs.gov/products

- A tool would perform the computations, document each data source (survey, table, variable names), and allow users to adjust the index to fit their needs, e.g., by adding or deleting variables.
- Sometimes, the indicators change from one year to another, making it challenging for each individual user to create the SVI from scratch.

Census Bureau data, especially ACS, undergo annual API code changes. When these changes occur, provide two options: one with the changes (the current way) and another that adjusts the data to facilitate comparisons year to year.

Geography Wish List.

Community Disaster **Resilience** *Zones (CDRZs):*

https://www.fema.gov/partnerships/community-disaster-resilience-zones

- "The Community Disaster Resilience Zones Act was signed into law by President Joe Biden on December 20, 2022. The Act amends the Robert T. Stafford Disaster Relief and Emergency Assistance Act. It requires FEMA to utilize a natural hazard risk assessment index to identify Census Bureau tracts which are most at risk from the effects of natural hazards and climate change."
- Researchers at TAMU, Duke University, and the University of Georgia are studying the application of CDRZs.

References.

Gu, D., Dillard, M., Gerst, M., & Loerzel, J. (2023). Validating commonly used indicators for community resilience measurement. *Natural Hazards Review*, *24*(2), https://www.nist.gov/publications/validating-commonly-used-indicators-community-resilience-measurement.

Forthcoming work by Michelle Meyer, Nathaneal Rosenheim, and others at Texas A&M University, Department of Landscape Architecture and Urban Planning and the Hazard Reduction and Recovery Center.

They replicated the CDRZ designation process using multiple SVI options. The interesting finding is that while the use of SVI significantly changed which Census Bureau tracts were selected, the different SVI options did not have a significant impact.

Federal Agencies are funding community resilience research. We have had discussions with the following people who expressed interest in learning more and collaborating with the Census Bureau.

- Jennifer Hegelson, Applied Economics Office, Engineering Laboratory, National Institute of Standards and Technology. Jennifer Hegelson@NIST.gov.
- Ellen Mecray, NOAA National Centers for Environmental Information, Eastern Region Climate Services Director. Ellen.L.Mecray@noaa.gov.
- Gina Eosco, NOAA, Division Chief for the Science, Technology, and Society Division at the Weather Program Office (WPO). gina.eosco@noaa.gov.

Appendix 3: Demonstration and Art of the Possible Use Case Summaries and Review of State Data Centers.

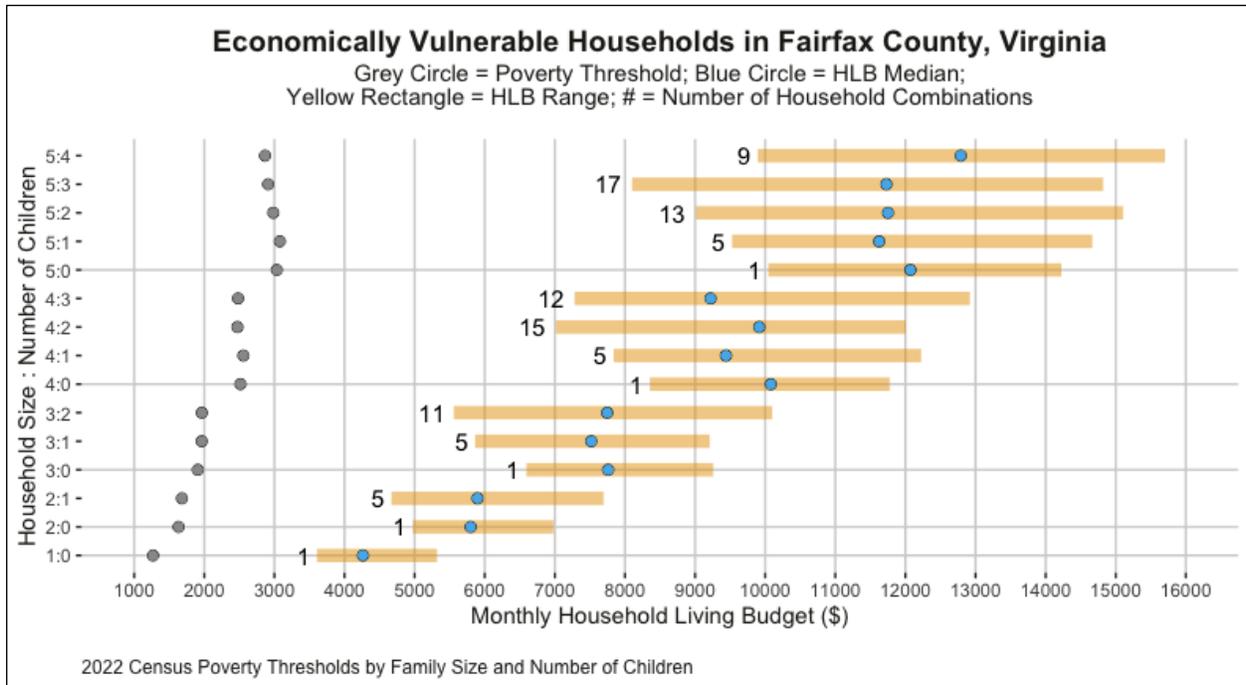**Demonstration use cases to identify CDE Capabilities to Develop Statistical Products.**

Summary.
Demonstration Use Case: The Importance of a Household Living Budget in the Context of Measuring Economic Vulnerability.

Purpose and Use: We constructed a Household Living Budget (HLB) to estimate the prevalence of economically vulnerable households at the Census Bureau tract level. In contrast, current measures of economic vulnerability, such as the Official Poverty Measure (OPM) and the Supplemental Poverty Measure (SPM), are at the national, state, and county levels.

The HLB evaluates the feasibility of creating a budget that captures a household's basic needs. This approach is being promoted as an alternate measure of poverty in the NASEM (2023) report on redrawing the poverty line. The HLB is constructed from many different data types, including administrative, opportunity, and procedural – how can they be integrated across different geographic levels, collected at different times, and for different analysis units? Some data sources are updated monthly, and some are updated every three years, which raises the question of how often a statistical product should be updated. The HLB includes an already computed transportation index. The question is what sort of metadata and benchmarking should be required for a statistical product like a transportation index to be housed in the CDE?

Exhibit A3-1 displays the HLB for households with one to five household members and the corresponding poverty threshold. The variation in the HLB includes both the variation between Census Bureau tracts and the variation between different household combinations with the same household size and number of children. Only when the household has no children is the HLB variation solely a function of geography. Family size and the age of children account for a large part of the variation in the HLB amount. We assume children under age 12 require childcare, which is expensive. A household defined as 5:3 has five members, including three children under 12, will require a much higher HLB because of childcare costs than a family with five members, including three teenagers. As one example, in Fairfax County, the yearly cost of childcare for an infant is $13,500, and this is for home-based childcare, which is the least expensive kind.

Exhibit A3-1. Monthly HLB descriptive statistics for various household sizes and number of children combinations in Fairfax County, Virginia. The number to the left of the range identified with a yellow rectangle is the number of household combinations included in the range.

CDE Criteria and Capabilities. We developed criteria to assess the feasibility of the Use Case. These criteria (left-hand column) were evaluated against this Household Living Budget Use Case (right-hand column in Exhibit A3-2). We constructed the HLB as our use case because it is a statistical product highlighting many of the capabilities needed in the CDE.

Exhibit A3-2: Criteria for Selecting & Implementing a Use Case. Does the Household Living Budget Use Case Meet these Criteria?

Initial Criteria for Use Case	How did the Use Case meet or not meet the criteria?
Assess relevancy of the research domain	• Developed a new poverty threshold based on a budget standard, proposed in An Updated Measure of Poverty: (Re)Drawing the Line on (NASEM 2023). • Addressed ethical concerns that current social benefit thresholds do not account for geographic price differences and therefore, benefit some more than others.
Determine availability of data from multiple sources across multiple frames	Discovered and used data sources to set adequacy standards. All component data sources are publicly available, they include: CMS Health Insurance Market Place for calculating health insurance and maximum-out-of-pocket costs.

Exhibit A3-2: Criteria for Selecting & Implementing a Use Case. Does the Household Living Budget Use Case Meet these Criteria?

Initial Criteria for Use Case	How did the Use Case meet or not meet the criteria?
	BLS National Compensation Survey percentage of employer contribution of health insurance Broadband cost data scraped from the internet DOT National Address Database to locate households within a Census Bureau block for scraping household broadband costs DOL Women's Bureau National data on childcare non-profit Feeding American Map the Meal Gap data to bring food costs down to the county level USDA Low-Cost Food Plan HUD Metro Area Advisory Fair Market Rent by Unit Bedrooms BLS Consumer Price Index Center for Neighborhood Technology Transit Affordability Index
Identify computing measurement requirements	Iterative proportional fitting to construct a synthetic population of households Computation of HLB components and overall HLB measure
Seek advice from Subject Matter Experts on data sources, the research approach, and implementation	US Census Bureau Geography Division, Enterprise Leadership Team, and others. University of Virginia Census Curated Data Enterprise team (2 former Census Bureau Directors, Communications Director, several well-known academics) NASEM, Connie Citro Duke University, Joe Hotz
Curate and document each step in the CDE process and describe outputs produced	All outputs, products, and references are included in the GitHub repository. Curation report: The Importance of a Household Living Budget in the Context of Measuring Economic Vulnerability: A Census Curated Data Enterprise Use Case Demonstration Application: The Cost-of- Living to Determine Food Insecurity. A repository: HLB Use Case GitHub Repository Data Tables and Derived Variables Data Code and instructions on running on the code Documents (literature, presentations, reports)
Assess processes and data sources with an ethical and fairness lens	The HLB resolves the one-size-fits-all social benefit thresholds by include geographic price differences and considering family composition and ensures transparency to creating a standard budget that is available to outside scrutiny.

Exhibit A3-2: Criteria for Selecting & Implementing a Use Case. Does the Household Living Budget Use Case Meet these Criteria?

Initial Criteria for Use Case	How did the Use Case meet or not meet the criteria?
Develop partnerships to access data from multiple types of organizations	Partnerships with Feeding America to bring food costs down to the county level. Center for Neighborhood Technology to get transit adequacy standards by household size and composition at the Census Tract level. BroadbandNow to get broadband adequacy standards at the Census Bureau tract level.
Viability of proposed platforms for interactive access to integrated data products for all interested in accessing the data while adhering to confidentiality and privacy rules.	All data are publicly available. The cost data to construct the HLB are a function of location and household combinations – not an individual.
Identify statistical product gaps and propose new data collection	• Data Gaps • Adequacy standards for childcare costs. Data are collected every 3 years by the Department of Labor and are based on state participation. In some cases, data were imputed due to lack of participation. • Other than childcare costs, this use case demonstrated that it would be unrealistic to fill in data gaps by conducting new surveys since data are needed at a small geographic level. Rather, it suggests that partnerships with for-profits and not-for-profits would need to be forged to resolve these data gaps.

Report: Lancaster V, Montalvo C, Salvo J, Shipp S, **The Importance of Household Living Budget in the Context of Measuring Economic Vulnerability: A Census Curated Data Enterprise Use Case Demonstration** Proceedings of the Biocomplexity Institute, Technical Report. TR# BI-2023-258. University of Virginia; 2023 October. DOI: https://doi.org/10.18130/p43z-c742.
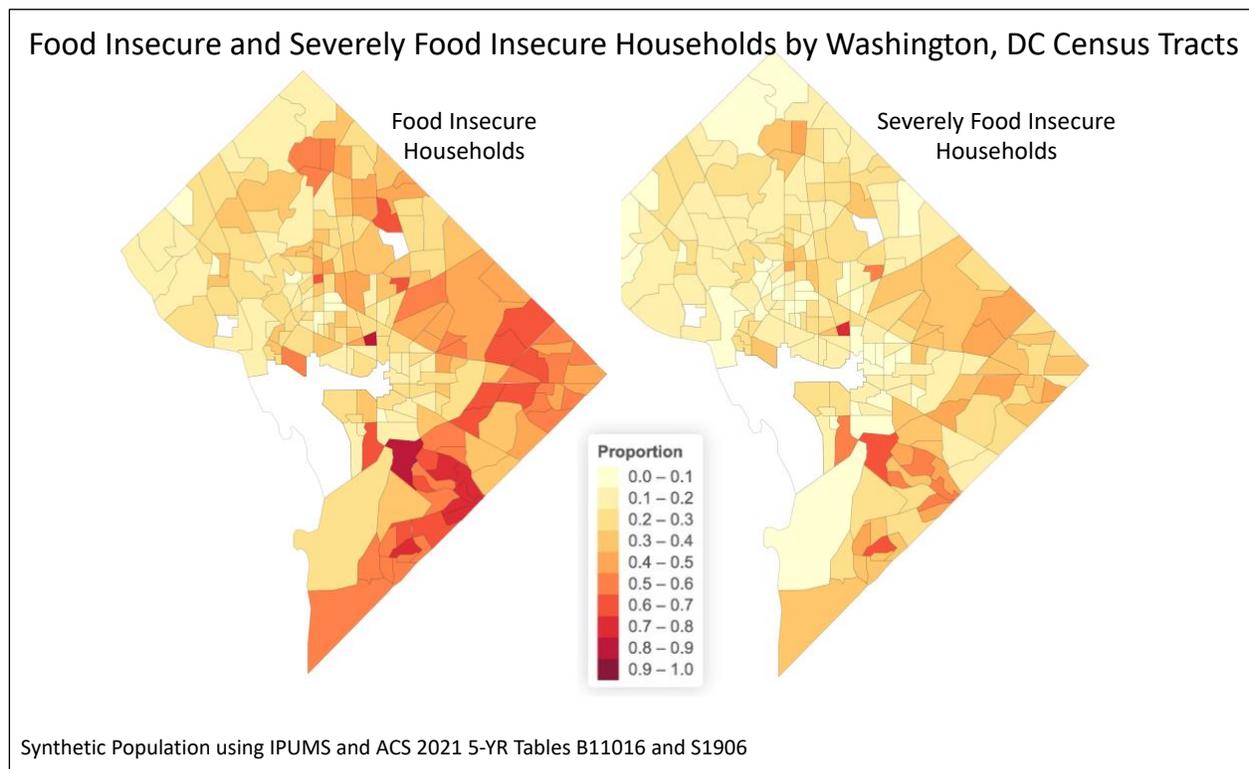
Journal: Lancaster, V., Montalvo, C., Salvo, J., & Shipp, S. (2024). Creating a Household Living Budget Using Publicly Available Data: A 21st Century Census Curated Data Enterprise Use Case. CHANCE, 37(1), 22-32. https://www.tandfonline.com/eprint/HGCKEAP73RIDAAX4U9BU/full?target=10.1080/09332480.2024.2324664

Demonstration Use Case 2: Applications of the Household Living Budget with a Focus on Food Insecurity.

Summary.
The USDA Current Population Survey Household Food Security Supplement Module (CPS-HFSSM) food insecurity measure has monitored food insecurity over time at national and state levels since 1995. But if a food insecurity measure is to inform action and target interventions, it must be constructed for smaller geographic levels that consider geographic price differences. The HLB is constructed at the Census Bureau tract level and incorporates three key determinants of food insecurity: household size and composition, household income, and food costs.

We provide an analysis of food insecurity in Washington DC households where food insecurity is measured using the HLB and the residual income method. The fact that our estimates of food insecure and severely food insecure households match those estimated using the HFSSM shows the promise of this method and the suitability of this research area for broader application across other geographic areas.

Exhibit A3-3. Prevalence of Food-Insecure and Severely Food-Insecure Households
by Washington, DC Census Tract using the residual method
based on the Household Living Budget.

By constructing the HLB at the Census Bureau tract level for all household combinations, we can identify tracts with a high prevalence of food-insecure and severely food-insecure households. Exhibit A3-3 displays the prevalence of the two food security categories at the tract level. Severe food insecurity is mainly concentrated in the eastern part of Washington, DC.

In the HLB Use Case Demonstration (Lancaster et al., 2023), the HLB was used to estimate economic vulnerability at the Census Tract level; this could also serve as a new measure of poverty. Finally, applications of the HLB for measuring economic insecurity measures and standards, food insecurity, housing affordability, and poverty would provide the opportunity to develop tools, methods, and processes to be applied elsewhere within the Curated Data Enterprise. For example, combining synthetic populations with the HLB estimates, administrative data on evictions, and social benefits would provide a wealth of information to inform policies that benefit low-income families.

Purposes and Uses. In this application of the HLB we construct a novel measure of food insecurity using the residual income method for Washington, DC, i.e., what is left over to spend on food after all other expenses? We then categorize households into three food secure categories: secure, insecure, and severely insecure. Our estimates of food-insecure and severely food-insecure households match those estimated using the USDA Household Food Security Survey Module, showing the promise of this method and the suitability of this research area for broader application across other geographic areas. Even though this application only focused on food insecurity, the other HLB components would benefit from this transparent approach.

Finally, applications of the HLB for measuring economic insecurity measures and standards, food insecurity, housing affordability, and poverty, would provide the opportunity to develop tools, methods, and processes to be applied elsewhere within the Curated Data Enterprise. For example, combining synthetic populations with the HLB estimates, and administrative data on evictions and social benefits would provide a wealth of information to inform policies better to benefit low-income families.

Report: Montalvo C, Lancaster V, Salvo J, Shipp S, **The Importance of Household Living Budget in the Context of Food Insecurity: A Census Curated Data Enterprise Use Case** Demonstration (2023) Proceedings of the Biocomplexity Institute, Technical Report.TR BI-2023-261. University of Virginia. https://doi.org/10.18130/2kgx-tv50.

Journal: Montalvo, C., Lancaster, V., Salvo, J. J., & Shipp, S. (2024). A novel approach to assessing food insecurity for small geographic areas using household living budgets. Food Security, 1-18. https://doi.org/10.1007/s12571-024-01494-5.

Art-of-the-Possible use cases to establish the feasibility of creating a statistical product.

The Net Undercount of Children Under 5 Years of Age in the Decennial Census: A Litmus Test for the Statistical Products First Approach.

Summary.
Purposes and Uses. The undercount of young children, ages 0-4, is a persistent problem in the decennial census and the largest net undercount of any age group (Exhibit 1). Not only was the undercount of young children very large, but the deficit of young children has increased relative to adults since 1980, reaching an all-time high in 2020. The undercount is disproportionately born by Black and Hispanic young children.

The Early Childhood Integrated Data Systems (ECIDS) bring together data sources on young children across multiple state-level agencies that serve young children. These state systems have allowed policymakers to respond more quickly and flexibly with up-to-date information about children and families, as exemplified in the COVID-19 pandemic with the closure of daycare and childcare centers. However, there needs to be a more robust disaggregation of ECIDS data to address questions of economic, racial, and gender disparities. This can be done through more extensive local, state, and

stakeholder data-sharing practices while working alongside community partners to highlight areas of need. Other data sources that could be included are Medicaid, TANF, SNAP, and school enrollment records, as state-level data could contain more granular information and resolve some of the limitations with federal data sources.

Integrating early childhood data can help develop a coherent approach towards early childhood education and services while revealing services that young children and their families access and addressing issues such as achievement gaps much earlier.[23] The degree of data integration at the state level varies substantially for various ECIDS. Early Impact Virginia (EIV) brings together data through the Early Childhood Foundation's ECIDS project while working with other stakeholders and policy specialists.[24] Initial funding was granted through the 2009 Statewide Longitudinal Data Systems (SLDS) as an initial step towards ECIDS and towards a centralized model for data development and integration.

A more centralized system allows for greater information linkages and accessibility, such as those found in North Carolina and Pennsylvania, where all data are consolidated in one database and linked as opposed to being deidentified beforehand. This model of ECIDS requires more time and poses more security concerns, but it only needs to be matched once and is much faster. Linkages and longitudinal data spanning multiple agencies are much harder to come by and must be generated, but this poses fewer security concerns. A hybrid model exists where linkages are only done once in a centralized system, using a data feed from individual agencies. There may be similar challenges in ensuring data matches up and reporting challenges, but it can operate off much fewer resources.

The ultimate utility of ECIDS for the purpose of improving estimates of young children is a relatively unknown right now; however, this approach holds promise provided that the needed research takes place. In a system aimed at filling-in the gaps that occur in surveys through the preparation of model-based estimates, the data provided from each individual source can be flawed so no single source will suffice. The value added occurs when data elements are combined to produce better model-based estimates, as with programmatic data providing counts of minority children underrepresented in surveys. Thus, the strength of this curation vision lies in the fact that different datasets contribute unique pieces to the overall estimation model.

Purposes and Uses.

Examples of purposes and uses that could be addressed with improved counts of young children include:

- Acquisition and distribution of funds for childcare and education.
- The calculation of rates for public health relies on denominators that define the population at-risk to illness and disease, as do determinations about the efficacy of an intervention to prevent and address such problems.
- Complete representation in a broad range of surveys

Examples of Purpose and Use questions include:

Baseline: What is the total child population in the state? What is the age, socio-economic, racial/ethnic, and geographic distribution of this population?

Research questions.

- Which groups of children and communities should be included in addressing gaps in childhood policies, planning, and programs?

---

[23] https://www.acf.hhs.gov/sites/default/files/documents/ecd/intergration_of_early_childhood_data_final.pdf
[24] https://rga.lis.virginia.gov/Published/2019/RD243/PDF

- **How many of these children are eligible for funded programs like subsidized child care, Reporting, etc., categorized by age and geography?**
  How many children are served by various combinations of early childhood services *This picture is needed before states can make informed decisions that typically affect many related programs.*
- What is the gap between children eligible for and receiving early childhood services? How do these access gaps vary by geography or community?

Based on our research and Listening Session, we identified five promising states to begin exploratory research – Pennsylvania, Minnesota, Georgia, Utah, and North Carolina.

We met with Third Sector Intelligence (3Si), a software firm specializing in the early childhood field. They have developed the **Child Universal Success Platform, CUSP**, as an Early Childhood Integrated Data System (ECIDS). The Gates Foundation funded 3Si to develop a public version of CUSP, called CUSP Public. CUSP Public has been vetted by five states representing about 20% of children ages 0 to 5% – Washington, California, Wyoming, Massachusetts, and Georgia.

Their CUSP Public process is to model the entire child population based on Census Bureau data. This public part contains use cases based on the Administration for Children and Families work with states and ECIDS active users. It includes indicators at the child level: child age, household income as a percentage of Federal Poverty Level and state median income, and parent labor force participation. Other data at the county level include child race and ethnicity, and eventually social vulnerability indexes, and language. CUSP Public has three main pages – child population, program eligibility, and scenario analysis. The private part of CUSP then integrates population data with state administrative data, configures CUSP based on state context, and creates a user interface and analytics platform.

Report: Salvo J, Lancaster V, Shipp S. **The Net Undercount of Children Under 5 Years of Age in the Decennial Census:** An Art of the Possible Use Case. Proceedings of the Biocomplexity Institute, Technical Report. TR# BI-2023-000. University of Virginia; 2023 October.
DOI: https://doi.org/10.18130/nzyj-m621.

Housing Affordability Art-of-the-Possible Use Case.

Summary.
Housing affordability makes for a logical Use Case. It is an important topic, as housing is widely acknowledged as a basic need, and the nation experiences a shortage of affordable housing. The issues surrounding housing affordability are complex. A confluence of factors has led to a shortage of affordable housing and many households experiencing high housing cost burdens. The factors are organized into three types: governance, shock events, and economics. Governance includes factors such as exclusionary zoning, which limits where high-density affordable housing can be built (for example, see Ikeda and Washington (2015).[25]

Other governance factors include *Not In My Back Yard* (NIMBY) sentiments and underfunding of housing assistance programs. Economic factors include wage stagnation and income inequality; local, state, and federal policy, including exclusionary zoning and the underfunding of housing assistance programs. Shock events such as the 2007-2008 Financial Crisis and the COVID-19 pandemic. Given this complexity, housing affordability is a topic that would benefit from the integration of multiple data sources, both survey and administrative data.

---

[25] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3211656

Housing affordability is highly complex and affected by many issues, including local, state, and federal policies, labor markets, monetary policy, domestic and international migration, economic conditions, and various forms of discrimination. The lack of affordable housing is tied to the inability to build wealth and homelessness. We limit our discussion to renter households; however, we also note that both renter and owner households can experience financial burdens due to housing costs along with many of the negative effects.

Exhibit A3-4 provides an overview of the pressures on housing affordability and the associated burdens. **A confluence of factors characterizes the modern housing affordability crisis.**

Exhibit A3-4: Contributing Factors to the Housing Affordability Crisis.

The factors are organized by three types: governance, shock events, and economics.

   Governance includes factors such as exclusionary zoning that has limited where high-density affordable housing ca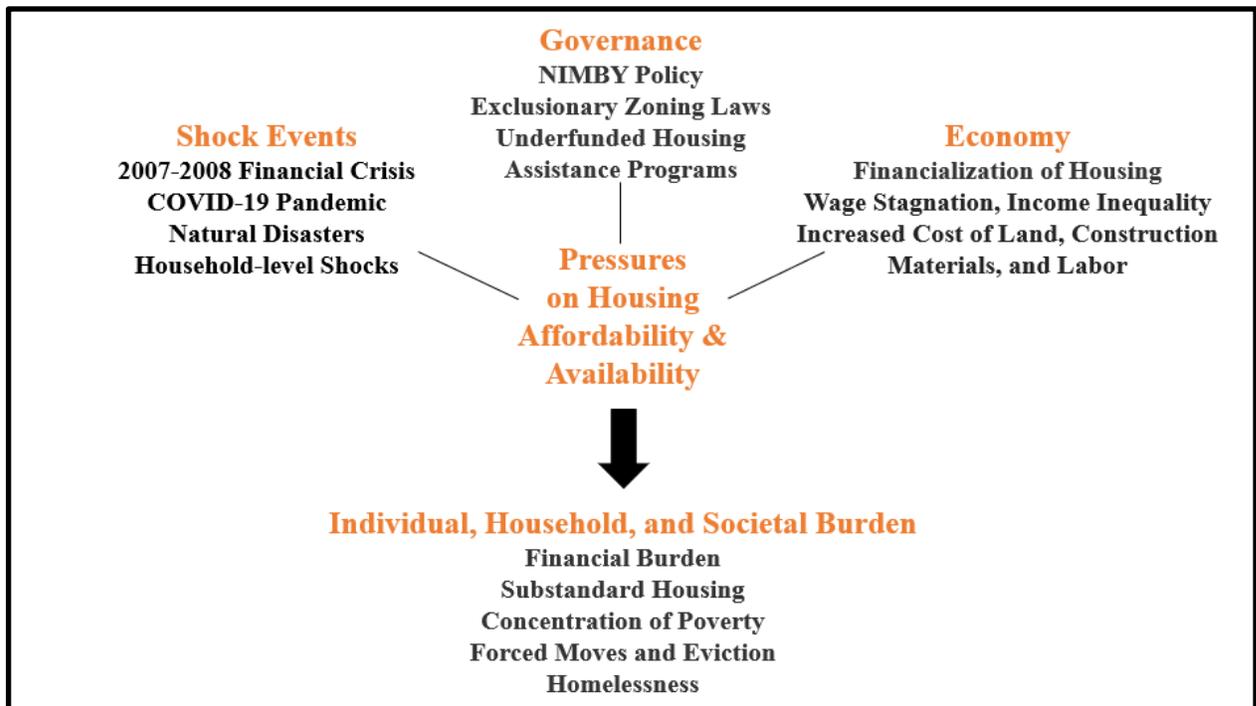n be built (for example, see Ikeda and Washington, 2015). Other governance factors include NIMBY sentiments and underfunding of housing assistance programs.

   Shock events have also impacted the affordable housing supply: Hurricane Katrina resulted in the destruction of a large amount of affordable housing in New Orleans (Herring and Rosenman, 2016), new house construction fell to historical lows in the wake of the 2007-2008 financial crisis (Goldman, 2009), and labor and material shortages due to COVID-19 pandemic have impacted housing starts over the last few years (Jefferson et al., 2022).

   Economic factors include the financialization of the housing stock, e.g., Airbnb, wage stagnation, and income inequality, which has driven up rents and incentivized developers to build higher-end houses (Ellen and Goodman, 2023). Other economic factors that contribute to the decrease in the availability of affordable housing are the increasing costs of land, construction materials, and labor.

Purposes and Uses. In our discussion of housing affordability issues, we identified four potential research areas that could benefit from integrating multiple survey and non-survey data sets to produce timely and granular statistical products. These demonstration use cases would provide the opportunity to develop capabilities (tools, methods, and processes) to be applied elsewhere within the Curated Data Enterprise.

> *Estimating Housing Cost Burdens.* What proportion of households are burdened by housing costs? How would a change in housing subsidies change this proportion? How would a change in local rents change this proportion?

> *Financialization of Housing and Multiple Property Ownership*. How many housing units are owned by private equity firms? What is the proportion of short-term rentals that are not primary residences? What are the effects of institutional investors and short-term rentals on house prices and rents? What are the effects of policies that aim to regulate short-term rentals?

> *Climate Change and Affordability.* Which areas are most vulnerable to increased energy bills because of climate change? How will that affect the proportion of housing residents burdened? How will climate-related disasters affect the current affordable housing stock? What are good locations for high-quality, affordable housing construction that is less vulnerable to the effects of climate change? Are these locations also in proximity to mass transit and other necessities?

> *Evictions.* How prevalent are informal evictions? How many people are at risk of eviction? How would economic (or other) crises affect evictions and those at risk of evictions? What was the effect of the Emergency Rental Assistance Program on evictions? Are there differences in eviction rates between different demographic groups after accounting for income and wealth?

> *Related to this topic*, we have been involved with and introduced the Census Bureau to the New America Foundation. The Foundation proposes creating a National Housing Loss Rate, like the unemployment rate, to raise the visibility of informal and formal housing losses by low-income families.[26]

Report: Wu E, Salvo J, Lancaster V, Stephanie S. **Housing Affordability – An Art of the Possible Use Case to Develop the 21st Century Census Curated Data Enterprise**, Technical Report BI-2023-262. Proceedings of the Biocomplexity Institute, University of Virginia; 2023 October. DOI: https://doi.org/10.18130/qgkd-va29.

State Data Centers.

This project focused on evaluating state constitutions and amendments, State Data Centers (SDCs), and representatives from the Federal-State Cooperative for Population Estimates (FSCPE). The goal was to understand better how state governments utilize data and to identify the federal, state, and private data sources they rely on. This report highlights the results of our data discovery process.

This project took an ambitious approach by exploring various methods to understand state and local data needs and how SDCs address those needs. We provide a list of near-term recommendations that the Census Bureau could implement to enhance the SDCs. One of our primary recommendations is for the Census Bureau to conduct a more in-depth analysis of a select number of states, starting with our

---

[26] https://www.newamerica.org/future-land-housing/blog/america-needs-a-national-housing-loss-rate/.

identified early adopters: Colorado, New Mexico, Kansas, and New York, as well as proactive data developers like Utah and North Carolina. This deeper dive could involve a thorough review of websites, interviews, focus groups with SDC and coordinating agency employees, as well as data users, and other methods to provide a more comprehensive overview of SDCs.

Report: Goswami T, Cai J, van der Geer M. Lancaster V, Kattampallil N, Shipp S, Molfino E. **A Review of State Data Centers Indicators, Tools, and Methods Technical Report,** 2024 Proceedings of the Biocomplexity Institute, Technical Report. University of Virginia. https://doi.org/10.18130/hbnk-ah54.

Appendix 4: Discussions with FSRDC Researchers and Experts.

FSRDC Researchers.

The US Census Bureau will need many Use Cases to develop a Curation System that aligns with the Curated Data Enterprise (CDE) framework. We brainstormed ways to develop these use cases efficiently and effectively. One idea was to build on researchers' work in the FSRDCs. We asked FSRDC researchers if they would participate in the Use Case program and what their incentives would be.

To develop Statistical Products, many use cases are needed to identify CDE capabilities to build. These use cases are focused on specific purposes and uses that can potentially be Statistical Products. FSRDC research generally addresses the other criteria for use cases:

- o Purposes and uses are defined that benefit both the researcher and the Census Bureau.
- o A variety of survey, administrative, and third-party data are used.
- o Statistical and computational methods are developed, leading to new capabilities.
- o The end products produce new theories, indicators, measures, and findings that could be translated to new Statistical Products.

Nate Ramsey, the Director of the Federal Statistical Research Data Center program in the Center for Enterprise Dissemination (CED), US Census Bureau, provided us with the names of FSRDC researchers to contact (see Exhibit A4-1).

We sent the following invitation to FSRDC researchers.

We are collaborating with the Census Bureau to develop a 21st Century Curated Data Enterprise[27] through use cases that integrate survey, administrative, and other data sources. Many research projects are excellent examples of use cases and are crucial for illuminating the capabilities to be built to produce statistical products. We would like your thoughts about whether researchers would participate in developing and curating use cases aligned with the CDE framework.

Our ask: A 45-60 minute discussion with our UVA Census Working Group in which we will:
- Provide a brief overview of the Curated Data Enterprise.
- A discussion of your research and how you document (curate) it.
- If asked, what would be your motivation to participate in developing a Use Case guided by the CDE?

Exhibit A4-1 lists the FSRDC researchers who accepted our invitation to discuss their potential role in curating their projects to inform and add to the CDE capabilities to be developed.

Overall, FSRDC researchers showed a positive reaction. There was general interest in CDE vision and implementation. Given resources and motivation, most would be willing to curate their research end-to-end. A universal request was that participation would help to streamline the Census Bureau approval process, formalize benefits, and more systematic use of FSRDC research by the Census Bureau. These FSRDC researchers noted that they share their research methods, tools, and findings formally in a memo and informally in conversations, but it was not clear if the Census Bureau used the information.

---

[27] Keller S, Prewitt K, Thompson J, Jost S, Barrett C, Nusser S, Salvo J, Shipp S. A 21st Century Census Curated Data Enterprise. A Bold New Approach to Create Official Statistics, Technical Report BI-2022-1115. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. DOI: https://doi.org/10.18130/r174-yk24.

Exhibit A4-1. FSRDC Researchers	Date
Jonathon Schroeder, National Historical Geographic Information System (NHGIS). Develops and applies methods of spatiotemporal analysis and visualization of Census Bureau data.	07/26/23
Ben Pugsley, Notre Dame. Studies, studies the behavior of young and small firms and their unique role in shaping aggregate economic conditions.	07/27/23
Florence Honere, University of Wisconsin. What determines who switches into entrepreneurship, either as a founder or early employee? How do prior work experiences influence start-up outcomes (i.e., survival, hiring, managerial practices, and alliances)?	08/23/23
Pargag Mahajan, University of Delaware. Studies the sources and consequences of labor mobility with a focus on immigrant entrepreneurship.	08/23/23
Hyunseob Kim, Federal Reserve Bank of Chicago. Examines the intersection of finance and labor economics.	09/01/23
Catherine Patterson, University of Chicago, Boot School of Business. Studies macroeconomics and labor economics, with a focus on how inequality across workers and firms can affect the economy's response to shocks.	No response

Selected thoughts from FSRDC researchers follow:
- o They owe their careers to Census Bureau data and would be happy to help the project simply out of goodwill.
- o Their research aligns with the CDE concept and there would be mutual benefit.
- o There was hope that the CDE framework would make linking data between sources much easier.
- o The data cleaning and linkage seen in the existing CDE use cases would be of major benefit to the Bureau.
- o One suggestion is that FSRDC researcher annual reports be used as an initial documentation source.

We also met with Lucia Foster and Randy Becker from the Center for Economic Studies. They discussed the two types of FSRDC research:
- o Research pull: The Census Bureau asks researchers to work with them to improve methods and data products or to create new data products.
- o Research push: FSRDC researchers apply to conduct a project.

They noted that the learning curve in FSRDC is challenging. Projects are generally about 5 years. Researchers hit their stride on their 2nd or 3rd project, somewhere between 6 to 15 years. Examples of how FSRDC research creates new knowledge are studies that provide feedback on quality issues with microdata, suggest new measures, and support the ability to combine Census Bureau microdata with data from other agencies Many of these ideas are posted in the Center for Economic Studies Technical Notes.[28]

---

[28] https://ideas.repec.org/s/cen/tnotes.html

Appendix 5. Stakeholder and Expert Meeting Summaries on Resilience and Related Topics.

We talked to various people to seek input or plan for a listening session on resilience and the economy, These conversations are summarized in this appendix. Appendix 6 provides a list and dates of other researchers, advocates, and nonprofit organizations we talked to.

Exhibit A5-1. Discussions with Resiliency Researchers, Advocates, and Agency Program Leads	Date
Experts at the Louisiana Resilience Conference	1/12-13/2024
Walter Gillis Peacock, Texas A&M University (TAMU) and Census Bureau	4/3/2024
Theresa Meyer, Texas A&M, the role of non-profits during disaster recovery	4/10/2024
John Van De Lindt and Jamie Kruse, co-directors, Center for Risk-Based Community Resilience Planning at Colorado State University	4/24/24
Gina Eosco, National Oceanic and Atmospheric Administration (NOAA), Weather Program Office (WPO), Science, Technology, and Society Division	5/20/24
Dustin Whitney	6/10/24

Summaries of selected interviews.

Louisiana Resilience Conference, an example of bringing the researcher community together to plan community action, January 11-12, 2024.

In January 2024, coastal experts from Louisiana and other states convened at the University of Louisiana Lafayette for a workshop titled "People in the Plan." This event was funded by the Restore the Mississippi River Delta Coalition in collaboration with the Coastal Protection and Restoration Authority (CPRA) and the Kathleen Babineaux Blanco Public Policy Center. The main objective was to explore how to better incorporate social considerations into Louisiana's Coastal Master Plan, which is crucial for addressing the ongoing coastal crisis affecting both coastal and inland communities.

The Coastal Master Plan, established by CPRA, is a stakeholder-informed and science-based framework designed to guide coastal restoration and protection efforts. Since its inception, particularly following the challenges posed by the 2005 hurricane season, CPRA has played a vital role in streamlining planning and implementation for coastal projects. This has resulted in Louisiana gaining national and international recognition for its best practices in coastal planning.

In preparation for the 2029 update of the Coastal Master Plan, the workshop brought together about 30 social scientists with diverse expertise in fields like land use planning, economics, data science, and climate resilience. [29] Participants aimed to identify relevant data and metrics that could promote fairness in the allocation of resources for risk reduction and restoration initiatives. Key discussions revolved around how to effectively acknowledge the intersection of social vulnerability and exposure to coastal risks, ensure that the distribution of resources benefits all community members, and enhance participation from marginalized groups.

Through various activities, the workshop participants evaluated existing data usage, metric selection processes, and funding priorities to identify gaps and opportunities for improvement. By considering

---

[29] Stephanie Shipp participated in this conference to learn about a state-wide initiative involving the community in metrics and data decisions. She introduced the Curated Data Enterprise framework and Statistical Products First approach.

these factors, the workshop aimed to promote coastal planning and secure better outcomes for Louisiana's communities.

Some of the questions explored during the workshop included:
- How are data sources used?
- Who chooses metrics, and which ones?
- What are the priorities?
- How do we include communities?
- How do we conceptualize predictive modeling?
- How do we build on and improve the most recent master plan?
- What have been the major complications in ongoing research/metrics/data? How do those factors impact the CMP?
- How is community well-being defined and considered?
- How do we define a coastal community? What about when the defining geography shifts (e.g., Census Bureau tract versus block versus place, cultural definitions, Tribal definitions, definitions of belonging to a coastal place, project type, and shifting coastal community composition)?

Short-Term Actions.
1. Historical Context Recap: Provide a detailed overview of past efforts in coastal master plans to inform communities about previous changes and outcomes.
2. Hindcasting Models: Implement models to show how coastal changes and impacts from major events (like hurricanes and oil spills) have affected regions over time.
3. Scenario Development: To help communities visualize different outcomes, create multiple scenarios that reflect potential future changes in demographics, industry, zoning, and stressors (like extreme weather).
4. Shorter Time Frames: Limit scenario planning to a maximum of ten years to align with decision-making processes and improve the usability of the predictions.
5. Coastwide and Regional Models: Develop both models to show broad and localized changes, enhancing decision-making support for communities.
6. Data Sharing Mechanisms: Identify metrics from various agencies to facilitate data sharing and funding opportunities across different coastal states.

Long-Term Actions.
1. Integration of Community Input: Develop a streamlined process for incorporating community feedback into future coastal planning and decision-making.
2. Capacity Building: Invest in training and resources for coastal communities and decision-makers to successfully navigate and implement coastal plans.

Aspirational Actions.
1. Collaborative Frameworks: Establish comprehensive partnerships among federal, state, and local agencies to collaborate on coastal management efforts.
2. Innovative Funding Models: Create and explore new funding mechanisms that encourage investment in sustainable coastal infrastructure and community resilience.
3. Vision for the Future: Foster a broader vision for how coastal regions adapt to changes over the next 20-50 years, considering ecological preservation and community needs.

These actions aim to enhance community involvement, provide valuable data insights, and promote long-term planning for sustainable coastal management.

In a follow-up meeting with Stephen Barnes, Executive Director, and Rachel Veron, LA First Director, we learned about the University of Louisiana Blanco Public Policy Center's role as a convener in this plan.

Walter Gillis Peacock, Texas A&M University (TAMU) and Census Bureau, April 3, 2024.

Professor, Dept. of Landscape Architecture & Urban Planning; Research Survey Statistician, Social, Economic, and Housing Statistics Division, Small Area Modeling & Development, US Census Bureau; Sandy and Bryan Mitchell Master Builder Endowed Chair; Sr. Faculty Fellow, Hazard Reduction and Recovery Center. https://www.arch.tamu.edu/staff/walter-gillis-peacock/

Suggestions for statistical products, tools, and research frames
- Explore developing products at finer resolutions (e.g., block group level) to facilitate county and local government-level planning. This may require modeling / small-area estimation for low-population areas.
- Create agreements with other agencies (e.g., the Department of Homeland Security (DHS), the National Oceanic and Atmospheric Administration (NOAA), and the Department of Agriculture (USDA)) to share data and improve the creation and/or validation of metrics and products.
- Create a research frame in Federal Statistical Research Data Centers (and possibly a public use file) that combines household, community, and impact data for natural disasters and hazards.

Discussion.
Professor Peacock emphasized the importance of a broad-based approach to risk management, focusing on mitigation (action to reduce impact), adaptation (learning), and recovery. He argued that these aspects are crucial in addressing vulnerabilities and preparing for future disasters. Professor Peacock also discussed the concept of risk, encompassing various impacts from business failures to household displacement and deaths. He highlighted the interplay between physical and social vulnerabilities, using the example of hurricanes that cause flooding and wind damage and how these impacts disproportionately affect vulnerable populations.

Professor Peacock discussed linking various data sources to better understand household vulnerabilities. He highlighted the ongoing work focusing on household data and the potential to supplement this with administrative data to reduce uncertainties. He emphasized linking this data to physical and social vulnerability elements. He suggested developing products to capture specific impacts and the likelihood of improvements. He uses the American Housing Survey data, LEHD Origin-Destination Employment Statistics (LODES), and OnTheMap (OTM) data.

He described the analysis as a Venn Diagram with the intersection of these components:
- Hazards (e.g., flood, wind, surge).
- Physical structures (e.g., building codes, roof design, garage doors, etc.).
- Social and community vulnerabilities.

Improving the alignment of various geographic maps, such as flood and surge maps, with Census Bureau geographies would help to understand better and address housing vulnerability and population exposure. Professor Peacock noted the need to develop reliable and valid measures for exposed populations. (A similar issue arose at the Louisiana Coastal Master Plan "People in the Plan" workshop in January 2024.)

Using high-resolution data at the block group level for community planning is critical because Census Tracts are too massive for many locations. Professor Peacock highlighted the challenges of working with

large counties and the need to shift focus from households to communities. He also mentioned the significance of considering social vulnerability and resilience when planning. He highlighted the lack of standardization in building codes and departments across different counties and regions, emphasizing the need for understanding these dynamics at a community level.

There is a pressing need to enhance community-level data to understand social vulnerabilities and facilitate planning at the local government level. He noted the need for more detailed and specific data beyond the household level to understand better physical vulnerabilities and linkages between social, economic, and physical factors. Professor Peacock also emphasized the potential value of data-sharing between different agencies, such as the DHS and the Census Bureau, to improve the accuracy and usefulness of the data.

Professor Peacock emphasized understanding events' broader social and economic impacts, particularly flooding and natural disasters. He suggested that incorporating data from various sources, such as DHS and NOAA, could improve the accuracy of models and predictions. There is a need for a centralized data frame, which would allow researchers to validate and refine their measures. He pointed out that this could facilitate better predictions and adaptations to future events, such as sea level rise and coastal storms. This work could be done by the Census Bureau, thereby reducing the piecemeal nature of current data collection efforts and individual researcher efforts.

Professor Peacock discussed large team science and **the potential use of engineering observatories**. One example is the National Science Foundation's (NSF) investment in design and collaboration with the University of Texas on a secure and computational environment project. Professor Peacock also mentioned their work with the NSF's resiliency modeling efforts, involving interdisciplinary teams of engineers and social scientists. He suggested that adapting existing models could make future projects more powerful.

Using open source and data for engineering models would enhance transparency and accessibility, which are crucial for validation and replication. These data would help validate models and acknowledge the importance of data biases. He suggested that thinking at different scales, from individuals to communities, could lead to more meaningful insights and identify new problems and issues. The concept of social capital is important to understanding connectivity among people.

The challenges and potential solutions in integrating various datasets for analysis were also discussed. He highlighted the difficulties in bringing together data types, such as economic and local government data, and the need for new data-sharing agreements to speed up and facilitate data access. He expressed frustration with limitations in accessing specific data, such as the National Flood Insurance Program data, which is crucial for studying events like Hurricane Ike in 2008.

Suggestions for others to talk to:

- Lori Peek, the director of the Hazard Center at the University of Colorado (on sabbatical).
- Texas Housers (did not respond to requests).

Michelle Meyer, Texas A&M University (TAMU), April 10, 2024.

Michelle Meyer is the Director of the Hazard Reduction and Recovery Center and an Associate Professor in the Landscape Architecture and Urban Planning Department at Texas A&M University. She received her Ph.D. from the Department of Sociology at Colorado State University (CSU) under Dr. Lori

Peek. She earned her BA in Sociology from Murray State University in Murray, KY, and an MA in Sociology from CSU.

Dr. Meyer's research interests include disaster resilience and mitigation, environmental sociology and community sustainability, and the interplay between environmental conditions and inequality. In particular, Dr. Meyer studies how disaster and environmental settings intersect with social structural forces that maintain or transform inequality. Her research projects include understanding the capacity of long-term recovery organizations, nonprofit efforts in resilience, disaster risk perception, social capital in disaster resilience, organizational energy conservation, volunteer training program evaluation, evaluation of disaster response plans for individuals with disabilities, social media use among vulnerable populations, how to increase protective action knowledge in Haiti, citizen science protocols for measuring storm-water condition fairness, and environmental attitudes and behaviors.

Discussion.
Dr. Meyer discussed the importance of using local data sources for community redevelopment and resilience, particularly in post-disaster recovery, and the role of non-profit organizations in assisting those in need.

Dr. Meyer discussed the integration of various datasets for community redevelopment and resilience, particularly in the context of post-disaster recovery. She emphasized the importance of using local data sources, including Census Bureau data, housing parcel, and non-profit data, to create more accurate and relevant resilience indicators. She also suggested defining catchment areas for non-profits to ensure their services are distributed fairly across communities. The challenges of tracking population displacement and recovery, particularly in low-income housing situations, were discussed, including the difficulty in matching property records over time due to a lag in the system.

Dr. Meyer described her work with long-term recovery groups, including non-profit organizations that typically emerge after disasters. These groups often include established charities like the United Way, Red Cross, St. Vincent de Paul, Catholic Charities, and various religious groups. The speaker noted that these groups often have more funding and resources available after large disasters, leading to criticism that they are "chasing disasters." However, these organizations are important in assisting those in need, particularly in areas with limited government and insurance support. For example, the Mennonites were explicitly mentioned for their work rebuilding homes for homeowners without insurance. The Methodists do case management. Catholic Charities oversees the financial aspects, especially tracking the money. She did note that these charities and non-profits are often "chasing disasters" as their grant money depends on assisting the community experiencing a disaster. This means they may leave one area before the work is finished to assist with the next disaster.

Dr. Meyer discussed the new case management model set up by FEMA, highlighting the role of case managers in matching funding opportunities to individuals' needs. She noted that nonprofits have traditionally been the primary providers of these services, but for-profit consulting companies are increasingly entering the field. The speaker expressed concerns about the profit orientation of these companies and their potential impact on the effectiveness of the case management system. She also suggested using open FEMA data to track case management at a more detailed level, similar to how FEMA funding is currently monitored.

Dr. Meyer noted the challenges immigrants face in accessing government aid, specifically through FEMA, and their need to rely on non-profit organizations as federal funding cannot be used to support them. She highlighted that non-profits often have to rely on non-federal funding sources to assist immigrants.

Suggestions for statistical products and tools
- Create a tool for communities to complete FEMA applications and for states and non-profits to apply for grants.
- Develop an infographic on calculating social vulnerability indexes, defining purpose, variables, and calculations.
- To measure recovery at the community level, provide more geographic granularity in the data below the Census Bureau tract level, e.g., block level and other geographies.
- Create a statistical product to provide information about the financial and organizational support provided after a disaster, including a metric for the percent of recovery achieved after a certain number of weeks or months.

Suggestions for Experts and Organizations to talk to:
- Nathanael Rosenheim, TAMU.[30]
  - Participated in the June 18 Listening Session and provided valuable insights.
- Jaimie Masterson, TAMU[31].
- National VOAD (Voluntary Organizations Active in Disaster) is an umbrella organization that brings together many disaster recovery partners.[32]

John Van De Lindt and Jamie Kruse, co-directors, Center for Risk-Based Community Resilience Planning at Colorado State University, April 24, 2024.

The Center for Risk-Based Community Resilience Planning at Colorado State University is a NIST-funded COE. The center's multidisciplinary work focuses on community resilience planning, materials science, and forensic science. The Center follows a collaborative and task-based approach involving economists, social scientists, and engineers from various universities. The center is considering forming a non-profit to sustain the center's work and create an open-source ecosystem.

The Center uses Census Bureau data to predict dislocation due to hurricanes, tornadoes, and other climate events based on race, ethnicity, and income at the household level. They then use a general equilibrium model to analyze community-level incomes and income and tax base changes. The goal is to provide information to communities to plan for and create recovery plans for all groups. Visualizing their community model includes physical infrastructure, economic models, and social services.

They use metrics like the Enhanced Fujita (EF) scale to determine the number of buildings that need to be retrofitted to withstand different levels of tornadoes. The team uses various models to predict the impact of natural disasters and the recovery process. The discussion also touched on the role of non-profits in disaster recovery efforts. (Note: EF assigns a 'rating' based on estimated wind speeds and related damage.)

IN-CORE Platform and Community Partnerships
The center develops community partnerships to create test beds for the **IN-CORE** platform, such as in Galveston, Joplin, Salt Lake City, and Monterey County. The use of **IN-CORE** varies by community and often involves collaboration with planning or resilience departments and emergency management.

More detailed Census Bureau data on multifamily buildings is needed to understand their impact on community recovery. Having individual apartment-level data would greatly aid in disaggregating and tracking the fate of households affected by events that displace people.

---

[30] https://www.arch.tamu.edu/staff/nathanael-proctor-rosenheim/
[31] https://www.arch.tamu.edu/staff/jaimie-hicks-masterson/
[32] https://www.nvoad.org/

Gina Eosco, National Oceanic and Atmospheric Administration (NOAA), Weather Program Office (WPO), Science, Technology, and Society Division, May 20, 2024.

Dr. Gina Eosco is the division chief of the Science, Technology, and Society Division in the WPO. She is a social scientist and risk communication expert who joined the NOAA Weather Program Office to lead the Forecasting a Continuum of Environmental Threats (FACETS) programs.

WPO is under NOAA's Office of Oceanic and Atmospheric Research (OAR). OAR supports research and works closely with the weather enterprise and NOAA's National Weather Service to develop, advance, and transition weather research to operations. Dr. Eosco's division supports extramural research. OAR also does intramural research. Both the external and internal programs are focused on measuring the societal benefits of NOAA's research.

At first, Dr. Eosco's division tried to take the same approach to research as the physical sciences. In 2019, they held a workshop and published the findings in 2020. The questions asked include.
> How to measure success in meeting NOAA's mission-critical factors, such as does NOAA's work:
> - Save lives?
> - Protect Property?
> - Enhance the national economy?

They concluded that to do this work, they would need to build an infrastructure for managing data for social science research.

Dr. Eosco discussed the **challenges and plans related to social science data at NOAA,** including integrating physical and social data for interdisciplinary research and developing a new flood inundation model. She also highlighted the need for better data accessibility, integration, and sharing within the organization and the ongoing data collection for severe weather events.

She highlighted the agency's efforts to improve the measurement of mission success, particularly in areas like saving lives and protecting property. She mentioned the recent **Bipartisan Infrastructure Law**, which provided funds for the "Societal Data Insights Initiative" to integrate physical and social science data. However, she noted the initiative's limited budget and the need to explore potential partnerships and data sources.

NOAA has an MOU with the Census Bureau to access data at the Census Tract level. Dr. Eosco noted that exploring the data requires a specific research question, but they would first like to explore it. This is not possible under the MOU. She also expressed concerns about the bureaucratic process required to access Census Bureau data, which they feel limits their ability to explore its potential uses. She also mentioned that the vast size of NOAA data exceeds the Census Bureau's capacity. *(Post-meeting note: Christa Jones, US Census Bureau's Chief of Staff, recommends that the relevant staff talk to NOAA about establishing a NOAA Research Data Center that would provide them access to Census Bureau microdata, which could be integrated with NOAA data.)*

Dr. Eosco highlighted the **interdisciplinary nature of their research questions and the challenges they faced in integrating Census Bureau data with their large-scale physical science data.** She expressed interest in using the public API to bring this data into their cloud environment. She also discussed **INSPIRE, "Integrated National Societal (Data) Platform for Interdisciplinary Research and Evaluation,"** which combines physical and social data to enable interdisciplinary research and evaluation. She highlighted the potential of their work to address societal issues, such as the impact of new products on society and the economic value of certain services.

Dr. Eosco introduces a **new flood inundation model that predicts areas at risk of flooding by considering precipitation, terrain, and river information.** The model is deterministic, but efforts are underway to develop probabilistic modeling. Questions arise about verifying the model's accuracy using mobility data or indicators of flood-related purchases. Broader concerns include understanding how public preparedness and protective actions vary by social vulnerability factors. Another area of interest is assessing impacts beyond mortality, such as on vulnerable housing like mobile homes.

Dr. Eosco discussed the **challenges and potential solutions for data integration and sharing within their organization.** She highlighted the need for better data accessibility and integration, particularly about vulnerability measures and weather-related data. She expressed concerns about the lack of large sample sizes and the need for more active funding. She introduced the "**Extreme Weather in Society Survey,**" conducted by the University of Oklahoma. The publicly available dataset measures public perceptions of their products and services.[33]

Dr. Eosco mentioned the **ongoing data collection for extreme weather events**, including tornadoes, that has been going on for over seven years. She said NOAA is developing corresponding surveys for tropical and winter weather and floods. She highlighted that 75% of the questions in each survey are consistent, while the remaining 25% are experimental, focusing on topics such as numeracy and product visualization. She emphasized the importance of longitudinal data analysis and the robustness of the sampling mechanism, which involves thousands of participants and the use of Census Bureau data.

Organizing Census CDE Listening Session for NOAA
Dr. Eosco is willing to host a Listening Session to highlight the Curated Data Enterprise and questions (purposes and uses).

Related Resources
Weather Program Office, Social Sciences, https://www.weather.gov/gjt/PeakToValleyWin15V2_SocialScience.
- Social Science https://wpo.noaa.gov/social-science/.
- Team: https://wpo.noaa.gov/team-page-2/.

Other stories and reports
- Open Data for All: The pitfalls and promises of making scientific results publicly available, https://wpo.noaa.gov/open-data-for-all/.
- Empowering Open Science: Unveiling WPO's System for Public Access to Research, Knowledge (SPARK), https://wpo.noaa.gov/empowering-open-science-unveiling-wpos-system-for-public-access-to-research-knowledge-spark/.
- Weather-Ready Nation Program, https://wpo.noaa.gov/weather-ready-nation-for-all-the-demographics-of-severe-weather-communication/.
- NSSL Research: Behavioral Insights, https://www.nssl.noaa.gov/research/social/.
- NOAA/SEES Supplemental Report: Social, Behavioral, and Economic Science Data in the National Weather Service. Author(s): Michaud, Michael S.; Trainor, Joseph E.; Philips, Brenda; League, Cedar; Corporate Authors(s): University of Delaware.; University of Massachusetts Amherst; Published Date : 2020, DOI: https://doi.org/10.25923/26er-dy91.

---

[33] Ripberger, Joseph; Krocak, Makenzie; Silva, Carol; Jenkins-Smith, Hank, 2021, "WX21", https://doi.org/10.7910/DVN/QYZLSO, Harvard Dataverse, V2.

Dustin Whitney, June 10, 2024.

Background.
"The Whitney Group, LLC provides executive search services to the financial services industry. The company offers services ranging from strategic senior hires to team build-outs for new products, services, and geographic coverage areas. It offers services in global banking and finance, global markets, asset management, real estate, insurance, competitive intelligence groups, and corporate officers and operations. The company was founded in 1983 and is based in New York City, with additional offices in North America, Europe, and the Asia Pacific. Whitney Group, LLC is a former subsidiary of Headway Corporate Resources, Inc." *The founder, Dustin Whitney, is very interested in federal statistics and regularly attends conferences such as the Population Association of America Applied Demography Conference.*

Highlights from the discussion.
Data Discrepancies in Occupation and Measures to Assess Housing Shortage
Interpreting household survey data compared to economic survey data, particularly concerning labor shortages, is complex. Discrepancies in data from the Bureau of Labor Statistics (BLS) and the Census Bureau, particularly in terms of occupation and industry classifications, make it challenging to match data from sources like Quarterly Workforce Indicators (QWI) and American Community Survey (ACS). Understanding the differences across surveys is necessary.

The mismatch between housing needs and the current housing stock, particularly in an aging society, requires improved measures that can be achieved through better alignment of household formation, inventory, and building permit data. This recent news piece helps show some of the challenges in the marketplace. See https://www.yahoo.com/news/census-data-shows-population-declines-233641119.html.

Improving Data Access and AI Integration.
It is essential to make statistical information more meaningful and accessible. However, recent changes in the presentation of this data have made it challenging for data users. AI has the potential to improve data accessibility and efficiency. Amazon Web Services has already deployed an AI module to this effect.

From a follow-up email: "There is a big market need for quality, streamlined government data that can help guide commercial planning. The need will grow in size and significance as the greater business and economic communities more thoroughly understand important demographic trends - working-age population growth stagnation putting pressure on the labor market, the implications of declining fertility rates, the real impacts of immigration playout, etc. I continue to see evidence that key strategic decisions are based on uncertain facts, which concerns me."

Appendix 6. List of Other Meetings: Researchers, Nonprofits, and Advocacy Organizations.

Exhibit A6-1. Researchers, Nonprofits, and Advocacy Organizations.

Researchers	Date
Alondra Nelson, Office of Management and Budget & Princeton University	04/04/23
Alisha Coleman-Jensen, Matt Rabbitt, Laura Hales, USDA Economic Research Service	04/06/23
Lars Vilhuber, Cornell University, Executive Director, Labor Dynamics Institute, presentation to team. Lars Vilhuber, is the American Economic Association's Data Editor, has reviewed more than 1,000 empirical articles since July 2019, and worked with authors to improve the reproducibility of their research compendia (replication packages). Curation lessons emerge from his work. (Vilhuber 2023).	05/11/23
Seema Iyer, University of Baltimore	06/1/23
David Dolson, former Director of Statistics Canada	07/18/23
George R. Carter III, Housing and Demographic Analysis Division, Office of Policy Development and Research, US Department of Housing and Urban Development	07/13/23
Ingrid Ellen, NYU Furman Center for Real Estate and Urban Policy	07/14/23
Whitney Airgood-Obrycki, Joint Center for Housing Studies of Harvard University	08/15/23
Joseph Hotz, Duke University	08/15/23
Stephen Barnes, Executive Director, and Rachel Veron, LA First Director, University of Louisiana, Blanco Public Policy Center	02/20/24
Andrew Aurand, National Low Income Housing Consortium DHRRC: advocates and researchers	5/20/24, 10/22/24, monthly check-ins
Yuliya Panfil, New America Foundation – Creation of a National Housing Loss Rate (like an Unemployment Rate)	5/15/24; 10/22/24, 12/18/24
Jennifer Hegelson, NIST, Community Resilience Program	9/6/2024
Tim Davidson and Regardt (Reggie) Ferreira, Tulane University's Disaster Resilience Leadership Academy and Consortium for Equitable Disaster Resilience (see Task 1, Prioritizing Purposes and Uses).	10/22/24 and 11/14/24
Nonprofits	
Joe Reso, President and Sammi Rowe, Research Systems and Economic Development Greater Fargo-Morehead Economic Dev Assoc.	02/07/23
Lamar Gardere, Executive Director, The Data Center (New Orleans), Allison Plyor, and Robert Habans	04/11/23; 10/10/23
Julie Tesch, President and CEO, Minnesota Center for Rural Policy and Development	
Cathy Lin, Director of Research and David Downey, President, International Downtown Association	05/25/23;07/21/23; 08/30/23
Third Sector Intelligence (3Si), Chris Strauz Clark, Ross Gilliland, David Halladin, Aradhana Roy), Will McCoy (Early Childhood Integrated Data Systems)	11/20/2023
Melissa Chiu, CNSTAT Director	6/20/23
Advocacy organizations (not including the Listening Sessions)	
Jennifer Bradley, Kresge Foundation	6/2/23; 10/18/23
Meeta Anand and staff, Leadership Conference-	10/18/2023
Loren Anthes, Head of Policy and Programs, Yuvo Health support community health centers	11/20/23
Count All Kids – Jasmine Jones and ~10 others	1/22/24
Gary Bauman, GDB Consulting, formerly Bauman Foundation	9/13/24
Chris Kingsley, Annie E Casey Foundation	9/18/24

Appendix 7: International Discussions –Synthesis of CNSTAT Talks and Agency Discussions.

International Statistical Agencies: What can we learn from other countries about how they are using administrative data to supplement, enhance, or create new data products?

Summary of CNSTAT Statistical Agency Presentations from Four Countries.
We present findings from four webinars held by the Committee on National Statistics for the US Census Bureau. Statistical agency officials from Canada, the United Kingdom, New Zealand, and Australia discussed modernizing their population census data collection and quality using administrative data.

Statistical Agencies worldwide are increasingly reliant on administrative data. Officials from Australia, Canada, New Zealand, and the United Kingdom discussed common challenges faced in modern censuses (and surveys), such as rising nonresponse rates and the need for accurate population counts at more granular geographic levels. The discussion highlighted how each country has implemented strategies to enhance survey response quality. For example, New Zealand integrated administrative data post-collection due to previous low response rates and the impact of Cyclone Gabrielle.

The officials also discussed the demand for more timely data, noting advancements in administrative tools across these nations to produce frequent outputs, such as the UK's midyear population estimates and New Zealand's experimental Administrative Populations Census. Furthermore, the operational efficiencies gained from using administrative data to optimize data collection processes are described.

The use of administrative data also brings about new data privacy concerns. While the timeliness of administrative data is seen as one of the main benefits of its use, this frequency makes it more difficult to ensure data remain confidential. The United Kingdom currently uses targeted method swapping census data as one of its disclosure control methods. However, this method is much harder to implement with administrative data because the frequency and quantity of data make it easier to reverse engineer the process. Currently, the United Kingdom is looking into alternative disclosure controls that can be used to replace this method.

Summary of Interviews with Four Statistical Agencies.

We also interviewed four statistical agencies in Australia, Canada, the United Kingdom (England and Wales), and Northern Ireland.[34]

Statistical agencies across various countries, including England/Wales, Canada, Northern Ireland, and Australia, increasingly leverage administrative data to enhance their census and survey efforts' accuracy, relevance, and timeliness. The UK Office for National Statistics (ONS) utilizes administrative data to combat declining statistical quality between censuses by generating admin-based population estimates (ABPEs) that provide population statistics months earlier than traditional methods. The ONS also explores using these data to study diverse topics such as migration, longitudinal life journeys, and demographic subgroups, often integrating it with surveys and commercial data for better insights.

Similarly, STAT Canada is replacing some survey questions with administrative data to improve non-response rates and the overall quality of its indicators related to Canadians' well-being. Northern Ireland employs administrative data for data linkage and targeted enumeration strategies, which improves census accuracy by mitigating issues like duplicate entries and tracking transient populations. The Australian

---

[34] We would like to thank Debra Stempowski, Assistant Director for Decennial Programs, US Census Bureau, who introduced us to these four statistical agencies, and Deirdre Bishop. Chief of the Geography Division at the US Census Bureau,

Bureau of Statistics (ABS) uses administrative data to refine census figures and adjust for missed entries by employing methodologies based on rental and occupancy data. These initiatives collectively illustrate a trend toward utilizing administrative data to enhance statistics and streamline future census processes.

Exhibit A8-1. Interviewees at Statistical Agencies.

Statistical Agency	Name and Title	Date
ABS	Anders Holmberg, Chief Methodologist, Methodology & Data Science Division	10 April 2023
ABS	Marcel van Kints, General Manager, Statistical Production and Digital Services Division	10 April 2023
ABS	Bruce Fraser, Director, Collection Management Unit	10 April 2023
ABS	Ross Watmuff, Statistician	10 April 2023
NI	David Marshall, Director of Census and Population Statistics	12 April 2023
CA	Stephane Dufour, Chief Statistician of Canada	17 April 2023
CA	Lynn Barr-Telford, Assistant Chief Statistician, Strategic Engagement	17 April 2023
ONS	Sonia Carrera, Deputy Director, Social Statistics Transformation, Analysis and Research	26 April 2023
ONS	Becky Tinsley, Deputy Director, ONS Local and Coherence	26 April 2023
ONS	Jennet Woolford, Director of Population Statistics Directorate	26 April 2023

ABS: Australian Bureau of Statistics.
NI: Northern Ireland Statistical and Research Agency.
CA: Statistics Canada.
ONS: UK Office of National Statistics.

More details are in this report:

Lanman K, Davis O, Shipp S. International Statistical Agencies: What can we learn from other countries about how they are using administrative data to supplement, enhance, or create new data products? Technical Report BI-2023. Proceedings of the Biocomplexity Institute, University of Virginia; 2023 October. https://doi.org/10.18130/2n54-sc22.