

# A 21st Century Census Curated Data Enterprise: *A Bold New Approach to Create Official Statistics*

Census Bureau Research and Methodology  
May 19, 2022

Team: Sallie Keller, Ken Prewitt, John  
Thompson, Steve Jost,  
Chris Barrett, Sarah Nusser,  
Joseph Salvo, Stephanie Shipp



**BIOCOMPLEXITY**  
INSTITUTE





“ The Census Bureau has innovated consistently over the decades. I am overjoyed to see the evidence of truly fundamental reforms to how the Bureau can supply statistical information to the country in ways that reflect 21st century approaches. ”

~Robert Groves



“ As we advance into the 21<sup>st</sup> century, we are experiencing increased demand for our data, struggling with challenges to traditional data collection methods, and exploring rich new datasources and tools that can revolutionize what we do and how we do it. Our success critically depends on our ability to seize the opportunities in front of us to deliver statistical products that address the increasingly complex and diverse needs of our users. ”

~Ron Jarmin



“ Thanks to Katherine Wallman, the 2000 Census replaced a misleading race/ethnicity question with one allowing Americans to ‘mark one or more.’ Census Bureau data are at their best when they fully reflect all of us, especially as the nation watched the steady growth of multiple race marriages and their offspring. The country is on the move, and so must its census. ”

~Ken Prewitt



“ This is a very exciting time to be working at the Census Bureau and to be part of the transformations further enabling the ways it is achieving its mission as a statistical agency. I am enthusiastic about how the disciplines of survey methodology, demographics, economics, data science, and computation are being brought together to measure the historic changes in our society and economy. ”

~John Thompson



# Purpose and Use



# Forces Driving Today's Innovations

- Enhancing the **gold standard** ... foundational to **ALL** official, state, and federal statistics
- Traditional data collection is unsustainable ... **cooperative response delivering diminishing returns**
- Meeting demand for **timelier, more granular, and robust data**
- **Protecting and securing** the information

Census Bureau must overcome these challenges:

- Be more **accurate and representative, stabilize cost, and minimize response burden**
- Respond to **conditions impacting racial and ethnic equity**
- Address **historical undercounts** across the country



The Bureau is **already exploring and implementing innovations** to modernize its operations:

- Making greater use of data the government already collects
- Expanding the use of third-party data and administrative records
- Linking data across all its surveys and censuses by unifying disconnected survey frames

# Introducing the Curated Data Enterprise

The U.S. Census Bureau has begun modernization activities towards creating a universal frame that could form the foundation for a transformative capability to integrate the Bureau's demographic and economic surveys and censuses with administrative and third-party data.

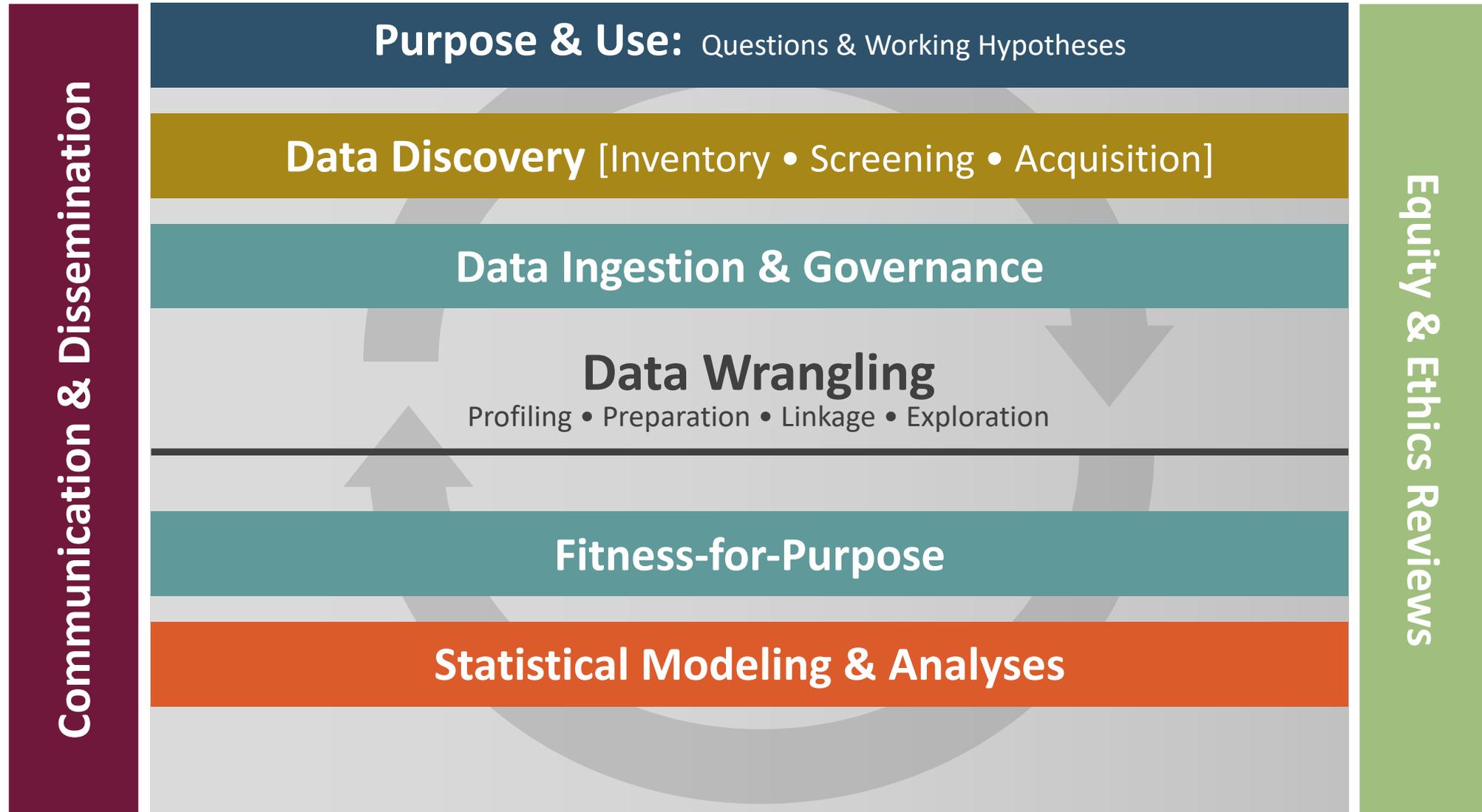
We describe this new, innovative state as a Curated Data Enterprise (CDE).

## CURATED DATA ENTERPRISE

*The Census Bureau  
of the Future*

- Both an infrastructure and a continuously evolving ambition to empower and enable Bureau scientists and their data users to develop new measures of people, places, and the economy.
- Exploits multiple data sources, from sample surveys, censuses, and other official administrative and third-party data, which can provide more robust, timely, and comprehensive measures when combined.

# Curated Data Enterprise Framework



# History: 2016

John Thompson asked  
JASON:

“ Starting with a blank  
sheet of paper, how  
would you design the  
2030 Census? ”



Organize around people vs housing units.



Start with “in-office” enumeration.



Develop a multifaceted strategy to find people  
who do not appear in the “in-office”  
enumeration.



Think beyond decennial, reconsider American  
Community Survey (ACS) and a **data  
ecosystem**.

*Alternative Futures for the Conduct  
of the 2030 Census*

<https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/alternative-futures-2030-census.pdf>

Contact: Dan McMorrow — [dncmorrow@mitre.org](mailto:dncmorrow@mitre.org)

November 2016

.JSR-16-Task-009

Approved for publication—distribution only by sponsor; U.S. Census Bureau, 4600  
Silver Hill Road Washington, DC 20233; Dr. John Thompson.



# History: One evening in December 2017



**Encouraged by John Thompson and Ron Jarmin,**  
Sallie Keller coordinated a meeting with:

- 5 former Census Bureau Directors
- Current Deputy Director, Census (ex-officio)
- Advocates
- Former federal officials
- University professors
- CNSTAT committee chair

At the end of the meeting, the group agreed to develop **an initiative to envision a new Census 2030** and **formed a working group.**

**Sallie Keller and Ken Prewitt agreed to co-lead**  
this Census 2030 initiative.

# History: Working Group (2018-2020)

## Co-Leads

## Census Bureau Directors

## Advocates

## Former Federal Officials

- **Sallie Keller**, Distinguished Professor of Biocomplexity, University of Virginia
- **Ken Prewitt**, Carnegie Professor of Public Affairs, Columbia University (Census Bureau Director, 1998-2001)
  
- **Vince Barabba**, Chairman and Co-founder of Market Insight Corporation, (Census Bureau director, 1979-1981)
- **Robert Groves**, Provost, Georgetown University (Census Bureau director, 2009-2012)
- **Steve Murdock**, Professor in Sociology, Rice University, (Census Bureau director, 2008-2009)
- **John Thompson**, Independent Consultant and University of Virginia, (Census Bureau director, 2013-2017)
- **Ron Jarmin** (ex-officio), Deputy Director of the Census Bureau
  
- **Arturo Vargas**, Executive Director, National Association of Latino Elected and Appointed Officials
- **Vanita Gupta**, President and Chief Executive Officer of the Leadership Conference on Civil and Human Rights
  
- **Cathie Woteki**, Iowa State University and University of Virginia, former Under Secretary for USDA's Research, Education, and Economics mission area
- **Stephanie Shipp**, University of Virginia, former SES at National Institutes of Standards and Technology

# History: An (Initial) Plan Emerges 2018

## BACKGROUND

Imagine the Decennial Census and American Community Survey counting the population more accurately, efficiently, at a lower cost, and taking advantage of 21<sup>st</sup> century advances in technologies and methods to integrate multiple sources of data.

## FUNDAMENTAL RESPONSIBILITY

Decennial and ACS (should) enlighten the public about the workings of a representative democracy, the goals of a free enterprise economy, and the realization of a socially just society.

## THE ISSUE

The time has come for a fundamental change in the conduct of the Decennial Census, and corresponding American Community Survey that can meet the growing operational constraints of the U.S. Census Bureau, declining response rates, increasing costs, and public trust.

## DECENNIAL CENSUS and ACS

The Decennial Census and ACS are more than apportionment and, by extension, redistricting. They also provide a national public good information platform.

**Anchoring on the Decennial and ACS, this thinking will evolve!**

# History: Sloan Foundation Support Late 2018-2021

- 1 **Vision evolves** to the framing of **the Curated Data Enterprise (CDE)**
- 2 **Formalized Advisory Panel** with former Census Bureau Directors, the Census Bureau deputy director, advocates, and former senior government leaders
- 3 **Held Advisory Group meetings** to advance vision
- 4 **Prepared white papers** to assess legal and constitutional challenges to a new approach for apportionment, redistricting, and allocation of federal funds.  
Conclusion: There are no barriers.
- 5 **Discussed ideas with decennial experts from industry and academia**  
Summary: Broad consensus and communications plan will be needed around this proposal to move forward
- 6 **Created University of Virginia-led CDE team**



# Today: Census Bureau MOU Tasking 2021 - 2022



## Identify and Characterize Stakeholder Communities

“ UVA will conduct workshops, webinars, and other activities with stakeholders to culminate into a comprehensive communication and outreach strategy relative to the innovation and modernization opportunities for the Census Bureau during this next decade.”

## Develop a Research Agenda

“ UVA will synthesize findings from workshops, webinars, and other activities with stakeholder communities to highlight big opportunity areas, technical challenges, and barriers to the innovation and modernization opportunities for the Census Bureau during this next decade.”

# Early Discussions with Experts

How will CDE **change stakeholder's approach** to their research and activities?

What are the **implications for state, local, and tribal governments** in their ability to identify and address needs and create informed policies?

What will be the **barriers for development and adoption**?

## Advice Received

- Rebuild trust with stakeholders.
- Partner and engage with local governments to build their capacity to use Bureau data products
- Provide data products to meet data users at all levels of expertise

**OMB:** Support a **decadal budget**.

**Congress:** Protect Bureau from **political interference**.

**Stakeholders:** Acknowledge there will be **trade-offs embedded in new strategies** and seek **input and cooperation** in making choices.

# Work Plan



**Convene stakeholder listening sessions to socialize the CDE concept and gather feedback**



## **Synthesize feedback**

- Formulate “use cases” that capture the art of the possible
- Capture concerns that will need research attention
- Hone communication and messaging
- Identify internal and external infrastructure needs



**Develop use cases to frame discussions with researchers for research agenda development**

- Demonstrate the utility and value of the CDE
- Illustrate research gaps and challenges

# Identifying and Characterizing Stakeholder Communities



## Congress and the Legislative Branch

Apportions the House of Representatives, allocates of funding to states, localities, and individuals, and formulates policy



## Research

Provides the foundation for economic, social, and policy research



## Public Policy

Supports public policy creation, program development, and administration including federal, tribal, state, and local governance



## Business and Commerce

Repackages the data to create "value-added" products and services and derives economic context through analysis to inform decision-making



## Non-Government Organizational (NGO)

Advocates for policies and programs and to monitor differential impacts across population subgroups and geographies



## Media

Informs the general public about the socio-economic situation of U.S. residents and changes in measurement



## General Public

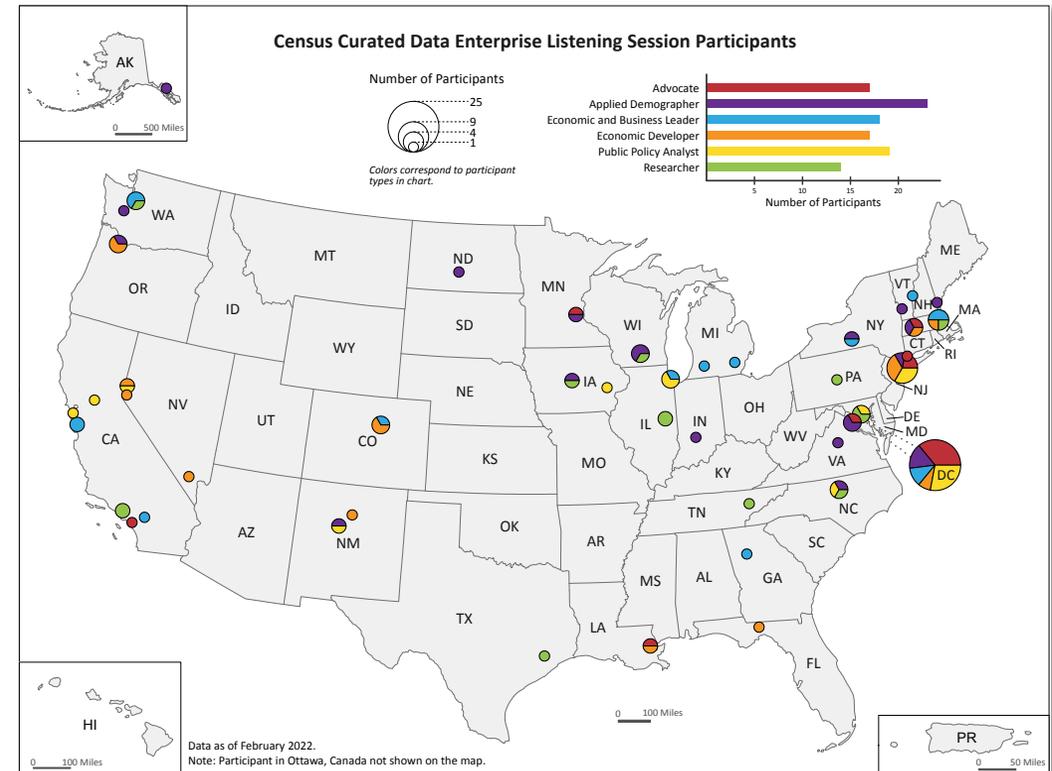
Data collection respondents and beneficiaries from every program and policy that is informed by Census products

# Listening Sessions with Stakeholders

Stakeholder Community	Date	Count
<b>Research Community</b>		
Researchers	Sep 1	7
Researchers	Sep 2	7
<b>Business and Commerce Community</b>		
Economic Developers	Oct 7	10
Economic Developers	Oct 8	7
Economists & Business Leaders	Oct 14	9
Economists & Business Leaders	Oct 15	9
<b>Non-Government Organizational Community</b>		
Advocates	Nov 4	8
Advocates	Nov 5	9
<b>Public Policy Community</b>		
Public Policy Analysts	Dec 2	8
Public Policy Analysts	Dec 3	12
Applied Demographers	Dec 16	13
Applied Demographers	Dec 17	11
<b>Total</b>	<b>12</b>	<b>110</b>

## 90-Minute Listening Sessions

- Describe the Curated Data Enterprise
- Share Census Bureau Integrated Frames Program
- Facilitate an open discussion about CDE over current state



# Listening Sessions: Questions We Asked

## Broad Discussion

What features of the economy and society would you **like to see measured** that the Federal Statistical System **is not measuring**, or is measuring inadequately?

What measurements need the **authority of Federal Statistics** to inform research and policy as a **trusted indicator**?

What data innovations are taking place in the private sector that **could be mirrored or leveraged** by Federal Statistics?

## Use Case Discussion

What data do you regularly consume?  
What data are **essential to you**?

How can data from the private and non-profit sectors **best be incorporated** into the CDE? What **partnerships** are needed?

What kinds of CDE outputs **would benefit** you the most?

# Early Take-Aways

## Overwhelming interest and encouragement to proceed

- Partnerships are key: private and other agencies
- Data products designed to meet user needs
- Information products that directly support other federal, state, and local agencies, and businesses
- More flexibility in using data
- Ideas for local, state, regional, and third-party data
- Internal and external communications
- Data access and confidentiality
- Sensitivities around use of administrative and third-party data

*“I'm impressed by this vision. Integrating geo with the other three frames is very powerful.*

*In this case [CDE], it's like one plus one equals 10. ”*

*Jack Dangermond, Esri  
Founder and President  
October 15, 2021*

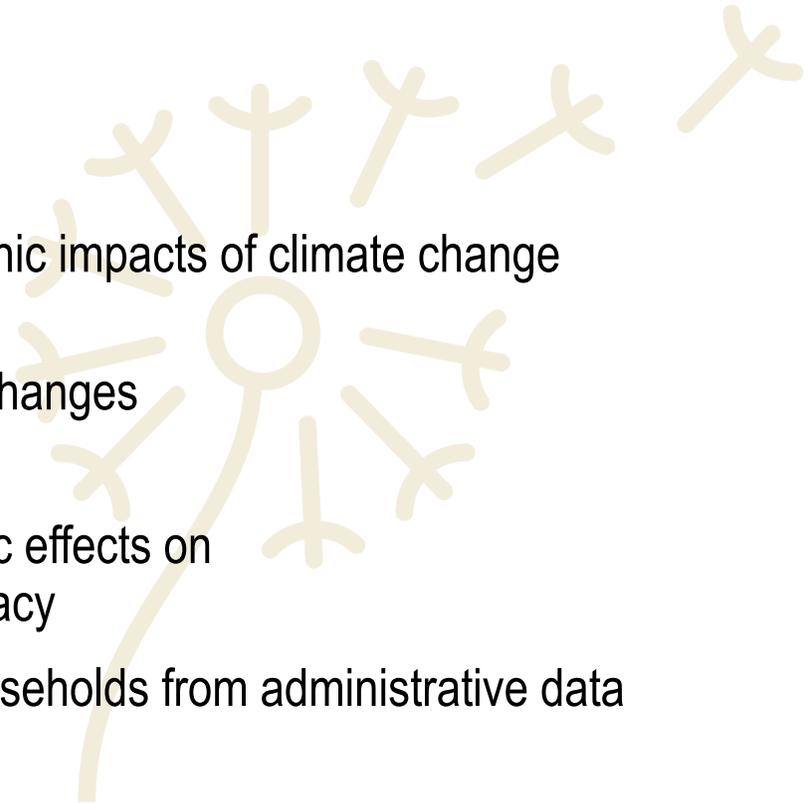
# Use Cases Will Play Key Role

Define and implement case studies to **illustrate the art of the possible, exercise research challenges, and find solutions:**

- Ability to address emerging topics
- Highlight use and development of new data collection, e.g., pulse type surveys
- Use of multiple data sources across multiple frames

Common suggestions (**wishes**) from listening sessions:

- **Nursing and residential care facilities**
- **Post pandemic gig economy**
- **Migration patterns**
- Housing, traditional and informal
- Counting children, child welfare
- Broadband access
- Rural and Tribal areas
- Developmentally disabled populations
- New race categories
- New construction data
- Economic and demographic impacts of climate change
- Income distribution
- Robotics and workforce changes
- Commuting patterns
- K-12 education, pandemic effects on outcomes and digital literacy
- Creating families and households from administrative data



# Early Research Topics Surfaced by Stakeholders

- Curation
- Data integration and linkage
- Trends with changing data sequences
- Integrating activities, geo and cyber locations, situational context
- Access and dissemination models for data and products
- Use of new technologies to collect data
- Timeliness / speed to market
- Expanding measures derived from ACS, addressing low sample density areas
- Multi-way partnerships and incentives to participate to include third-party data, other agencies , state and local governments
- Privacy and confidentiality
- Trust in collection, operations, and data products
- Respondent opportunities and reducing burden

# Early Research Topics Surfaced by Stakeholders

- **Curation**
- **Data integration and linkage**
- Trends with changing data sequences
- Integrating activities, geo and cyber locations, situational context
- **Access and dissemination models for data and products**
- Use of new technologies to collect data
- Timeliness / speed to market
- Expanding measures derived from ACS, addressing low sample density areas
- Multi-way partnerships and incentives to participate to include third-party data, other agencies , state and local governments
- Privacy and confidentiality
- Trust in collection, operations, and data products
- Respondent opportunities and reducing burden

# Research Roadmap Development for Initial Research Topics



**Convene stakeholder research consultation sessions**



**Synthesize research brainstorming**



**Develop 12-18 month research agenda for characterizing and developing capabilities of CDE**

- What is **known** about the 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 challenges?
- Which **researcher might be missed** when thinking about the CDE?
- Which aspects of the research is the Census Bureau well positioned to **execute internally**? To **partner** through COAs or other mechanisms?

# Initial Research Topics for Roadmap: Case Studies

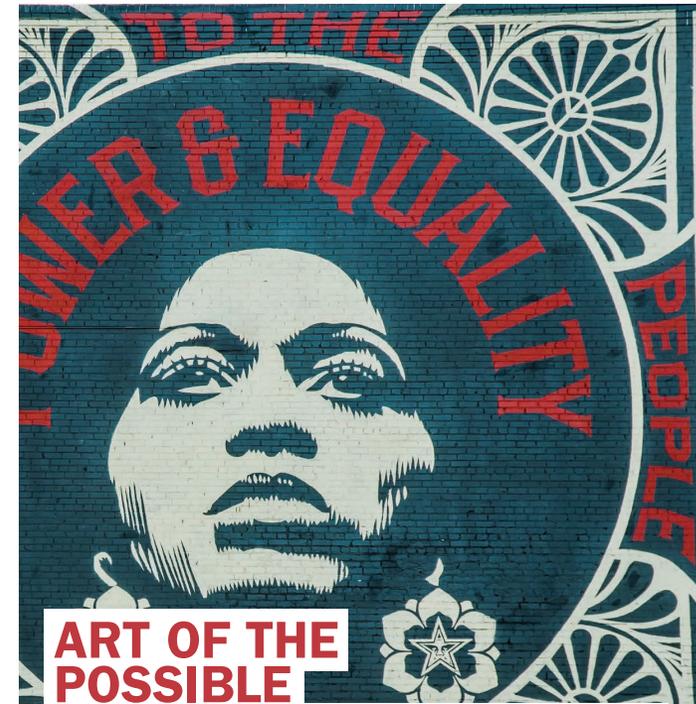
What **criteria could be used for selecting** a Case Study to support CDE development?

How could a Case Study **exercise some aspect** of the CDE?

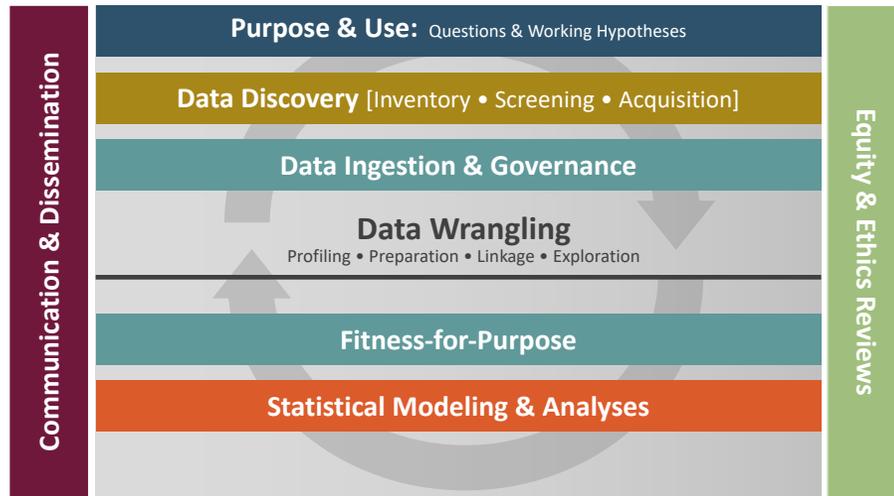
Potential required components to assess in the selection of a case study:

- The relevancy of the research domain and the **identification of purposes and uses**
- Value and **feasibility of the CDE** approach
- **Availability** of survey, administrative, opportunity, and procedural data
- Ability to identify **measurement requirements**
  - Units of analysis and their characteristics
  - Temporal sequences
  - Geographic detail
- **Methods** for imputations, projections, statistical analyses
- **Partnerships** for data access
- **Experts** to advise

What are the **curation requirements**?



# Initial Research Topics for Roadmap: Curation



## Discussion Topics:

- What are the general principles for developing data standards? How might they be applied to curation of the CDE Framework components?
  - How do purpose and use capture facilitate (or complicate) curation processes?
- 
- What is well known that should be maintained, (e.g., metadata and paradata capture for survey data elements)? How do these concepts apply in the context of non-survey data? What needs to be added?
  - How does quality evaluation work in the CDE, particularly given every data element will have the opportunity to have its source of information?
  - For which components of the Curation Framework is automation feasible?



# Initial Research Topics for Roadmap: Data Linkage & Integration

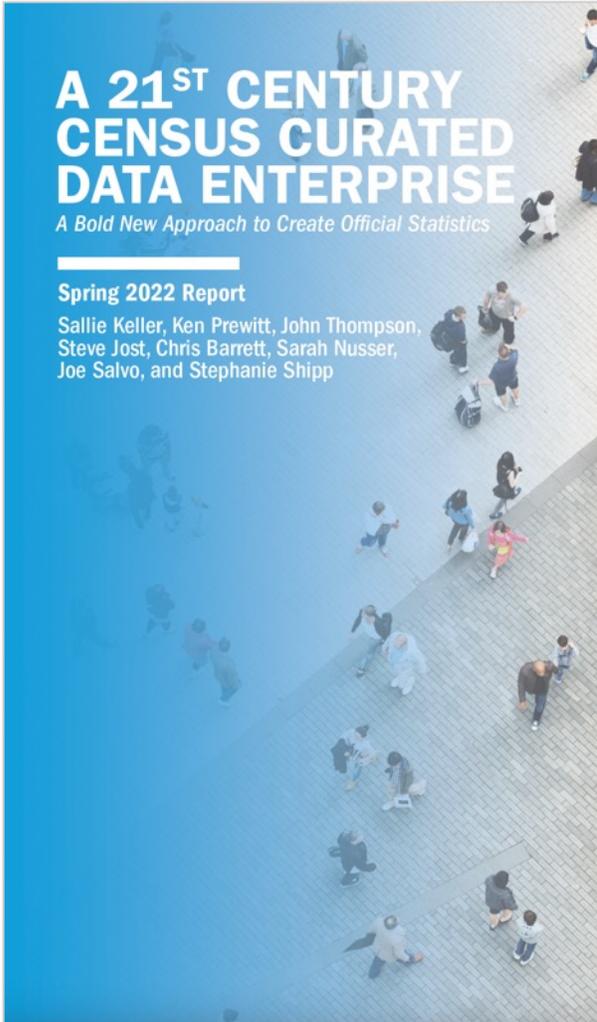
## Discussion Topics:

- **Data types** to consider are survey, administrative, opportunity, procedural, third-party, state, tribal, and local data.
- **Risk mitigation** associated with data source availability?
- Does a focus on purpose and use **change how to think** about data integration and linkage?
- What are **strengths and known limitations** of each of these sources?
- In addition to sampling errors such as coverage and bias, what components should be included in **quality metrics for new data products**? Analogues to total survey error?
- Stakeholders express needs for **user-defined geographies and classifications**. What research should be initiated to meet this important demand?
- The Census Bureau is **committed to increasing diversity, equity, and inclusion** in its business processes and in data products provided to stakeholders. How can linking and integrating multiple data sources help **meet this goal**?

# Timeline: September 2021 to August 2022



<https://biocomplexity.virginia.edu/institute/divisions/social-and-decision-analytics/census>



**Spring 2022 Report**

Sallie Keller, Ken Prewitt, John Thompson, Steve Jost, Chris Barrett, Sarah Nusser, Joe Salvo, and Stephanie Shipp



The screenshot shows the website's header with the UVA Biocomplexity Institute logo on the left and navigation links (THE INSTITUTE, OUR APPROACH, OUR RESEARCH, UVA: OUR SHARED VISION) on the right. A large banner with a colorful, crumpled paper background features the title 'A 21st Century Census Curated Data Enterprise'. Below the banner is a section titled 'Supporting Innovation at the U.S. Census Bureau' with a Latin placeholder text. A grid of six content cards follows: 'Reports' (Spring 2022 report), 'White Papers' (New ideas for a new era), 'Appendices', 'In the News', 'Innovation' (U.S. Census Bureau's History of Innovation), and 'Use Cases' (Art of the Possible).

**Supporting Innovation at the U.S. Census Bureau**

Numquam sagittis voluptate! Quidem diam quasi? Donec turpis dis aspernatur, itaque libero! Nunc quas imperdiet, maxime! Dignissimos blandit



# A 21st Century Census Curated Data Enterprise *What are the opportunities?*

THOUGHTS?

