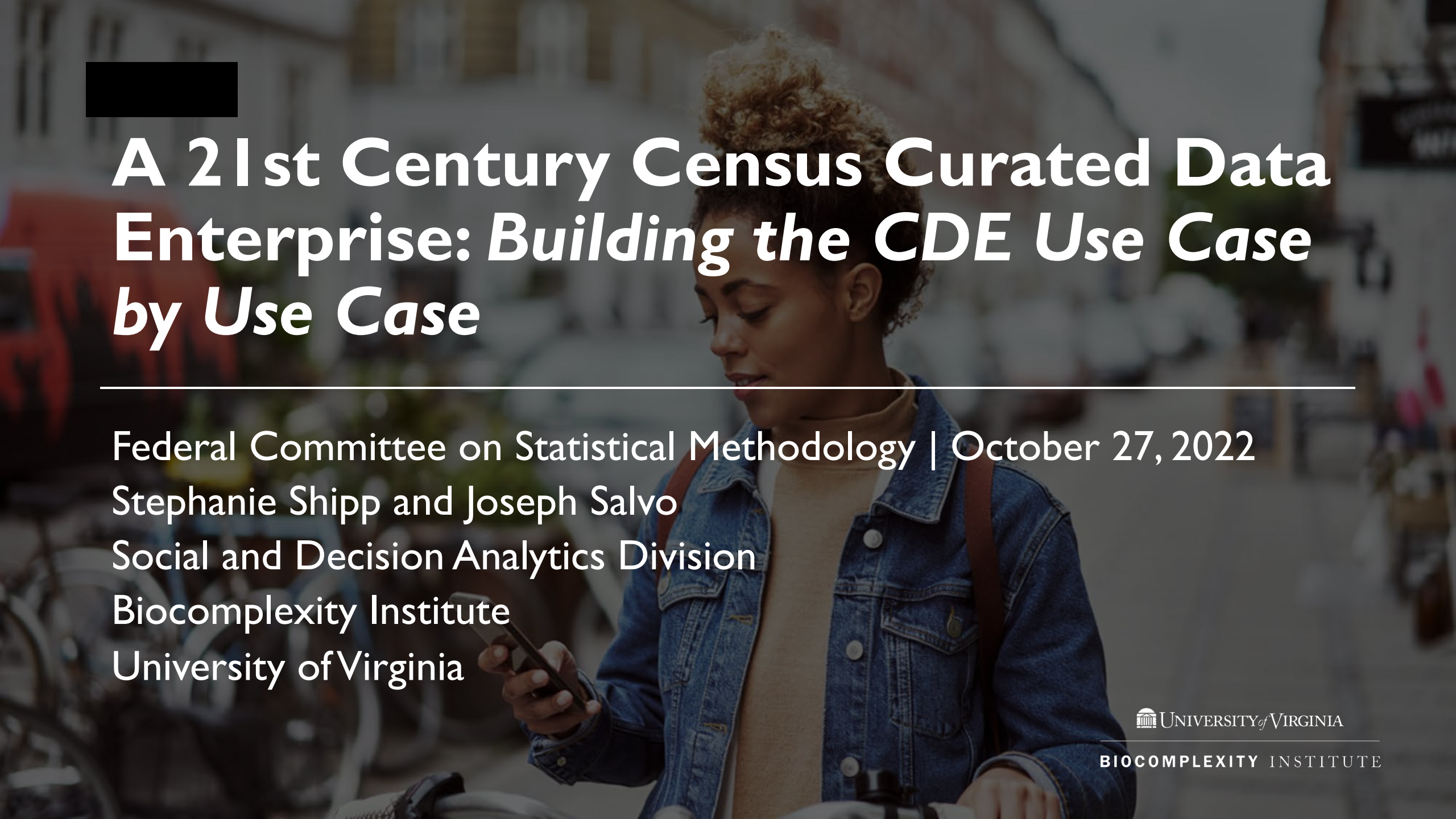


A 21st Century Census Curated Data Enterprise: *Building the CDE Use Case by Use Case*

Federal Committee on Statistical Methodology | October 27, 2022

Stephanie Shipp and Joseph Salvo
Social and Decision Analytics Division
Biocomplexity Institute
University of Virginia

UNIVERSITY of VIRGINIA
BIOCOMPLEXITY INSTITUTE

Paradigm Shift for Census Bureau: Focus on Purpose and Use

With the accelerating pace of change in our society and economy, we must constantly innovate how we collect and report official statistics to understand who we are, and where we are going.


Equity and Inclusion


Counting Children


Family Units


Rural Communities


Local Businesses


Tribal Areas


Vulnerable Populations


Elderly Residential Care


The Unhoused


Migration


Climate Change


Gig Economy


Workforce Automation

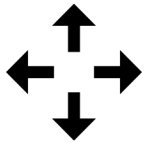

Differential Undercount


Broadband Deployment


Emergency Preparedness


Economic Mobility


Commuting Patterns


Changing Demographics

Introducing the Curated Data Enterprise

"**Linking across time, and surveys, supplemented with third-party and administrative** data can help improve the coverage and completeness of Census Bureau data, innovating **beyond a survey-alone model** that has reached its scientific and practical limits." *(Keller et al. 2022)*

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.

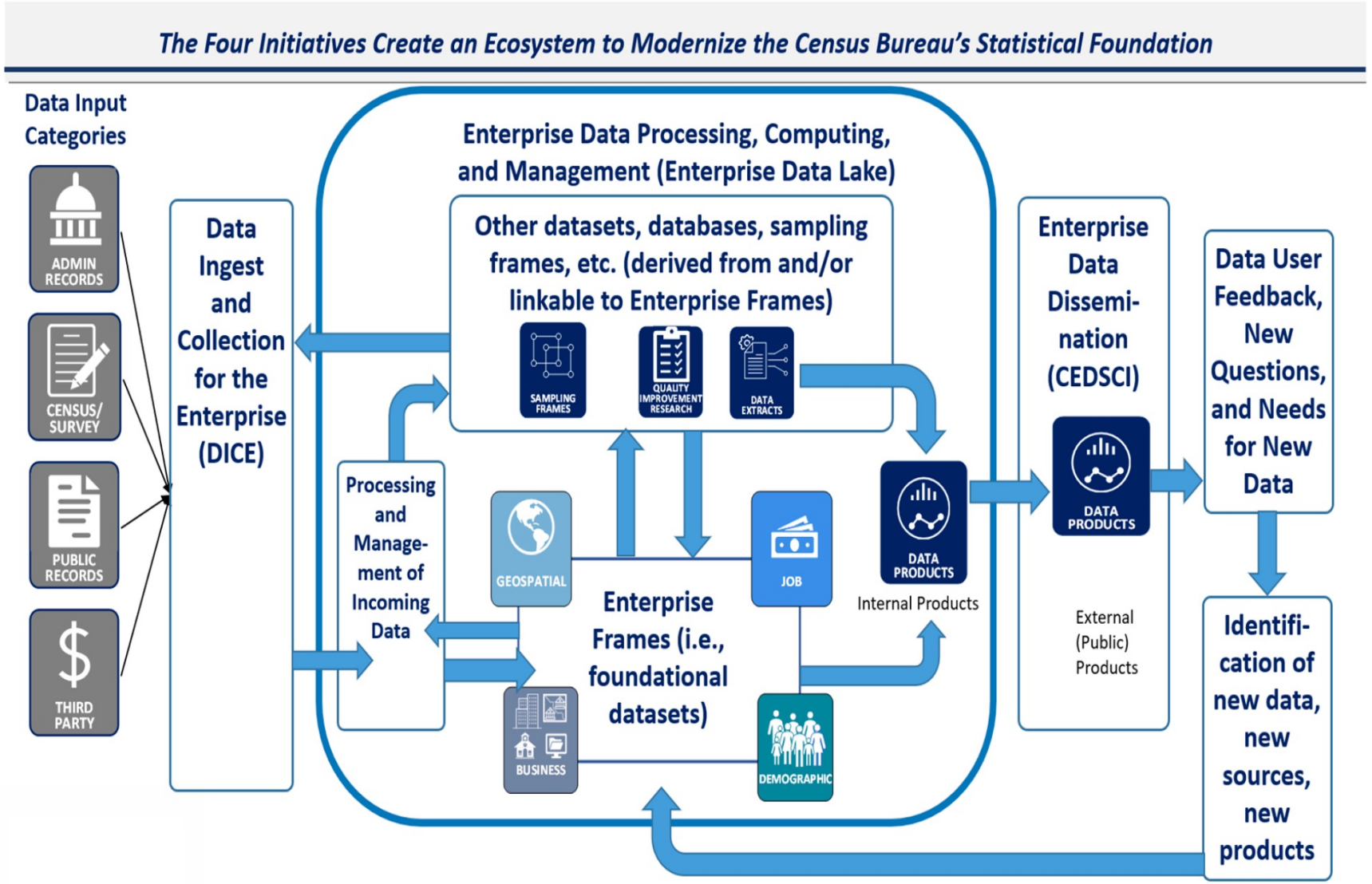
Exploits multiple data sources, from sample surveys, censuses, and other official administrative and third-party data, which can provide more robust, timely, and comprehensive measures when combined.

Building a Census Bureau Modern Data-Centric Ecosystem

Change is here!

"Our changing culture and rapid changes in data and technology tell us that **censuses and surveys alone, while still critical, can no longer answer society's questions completely** or quickly enough to satisfy the modern appetite for information."

Michael Thieme, Oct. 5, 2022



Source: https://www.census.gov/newsroom/blogs/research-matters/2022/10/technology-transformation.html?utm_medium=email&utm_source=govdelivery

History from 2015-2022


Alternative Futures for the Conduct of the 2030 Census

November 2016

JSR-16-Task-009

JASON

2015-2016

John Thompson asked JASON:

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

2017

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

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

2018-2020

Census 2030 committee concluded that:

"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."

ALFRED P. SLOAN FOUNDATION

2021-2022

Census Bureau Tasking

- Identify and Characterize Stakeholder Communities
- Develop Research Agenda

Thinking evolved

- Initially proposed essential data only decennial
- Decennial Census and ACS are more than apportionment and redistricting
- They provide a national public good information platform
- Proposed 21st Century Curated Data Enterprise
- Sloan Foundation and Census Bureau funding

The Curated Data Enterprise Team

Steve Jost

John Thompson

Sallie Keller

Ken Prewitt

Matthew Snipp

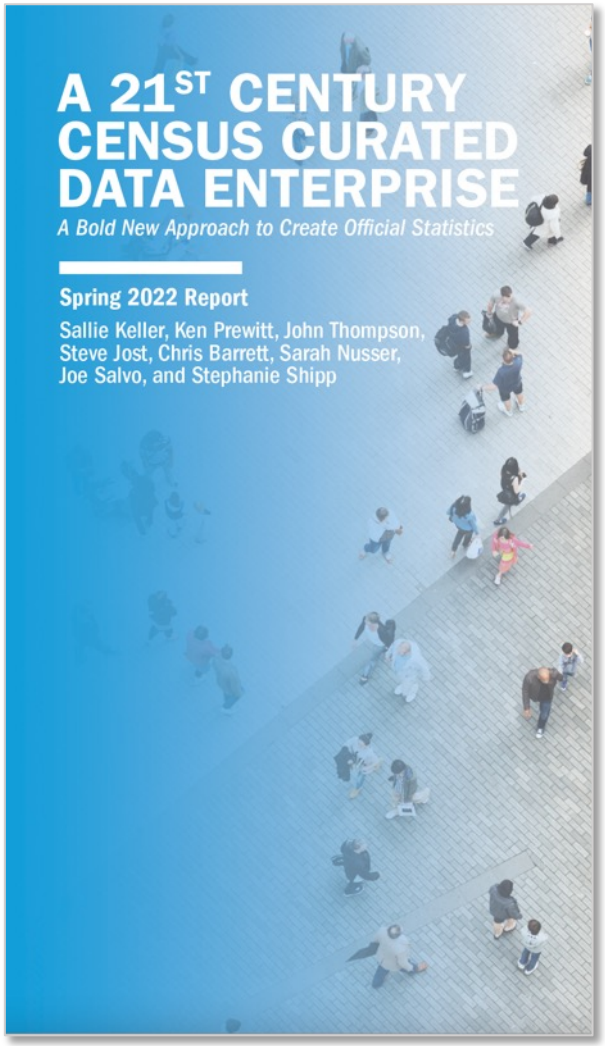
Chris Barrett

Sarah Nusser

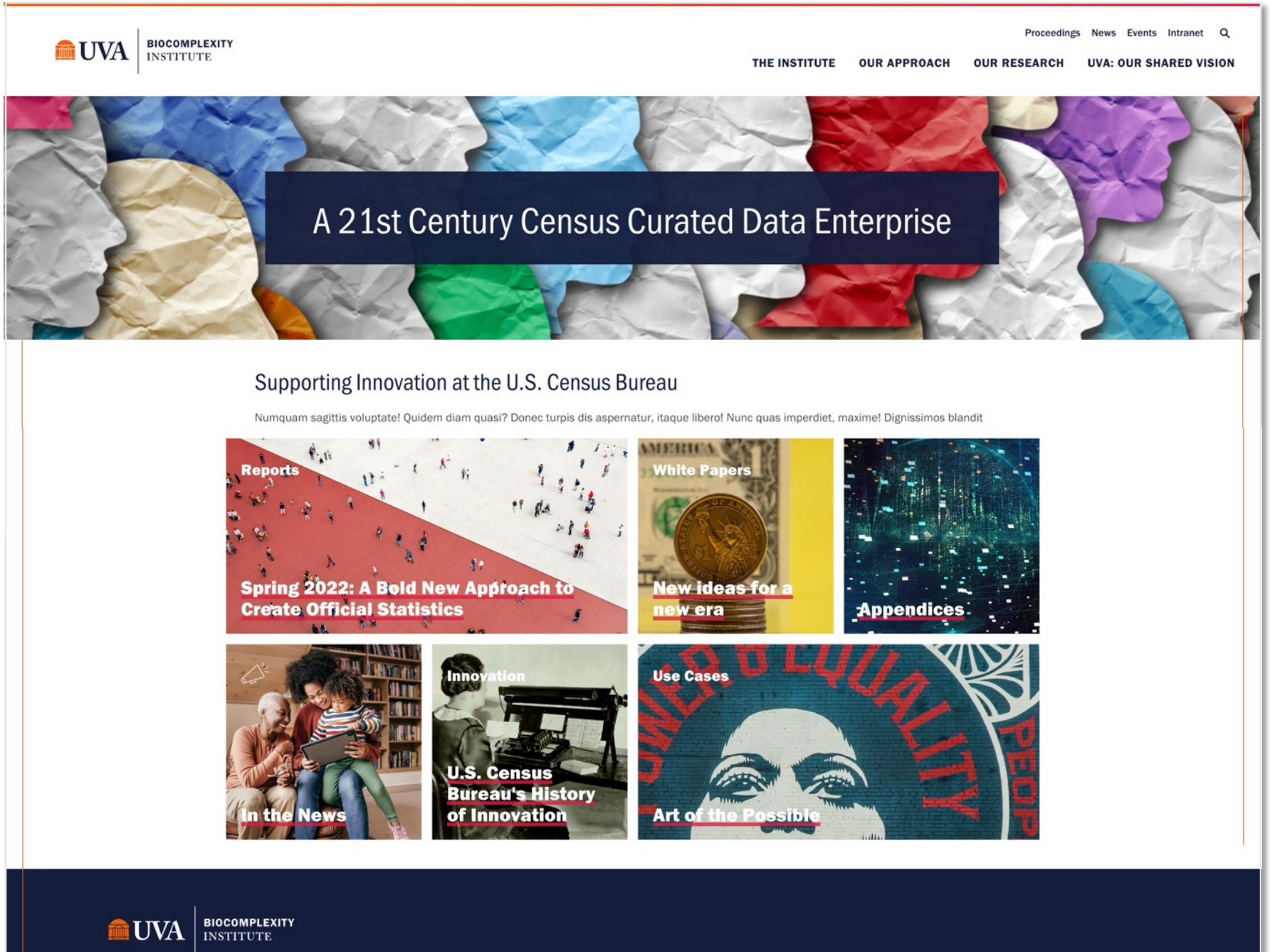
Joe Salvo

Stephanie Shipp

Zhengyuan Zhu

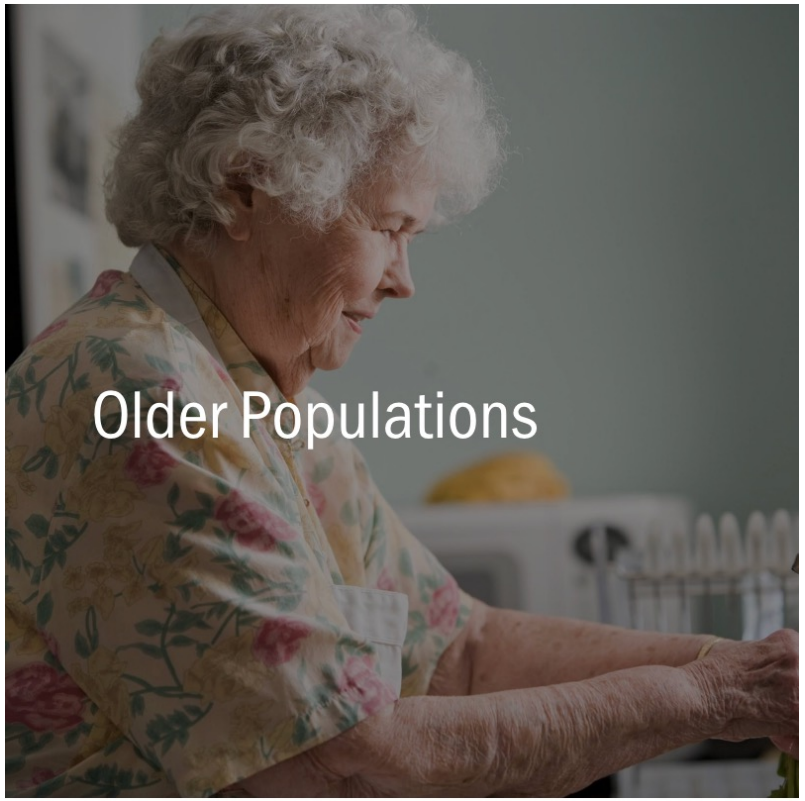


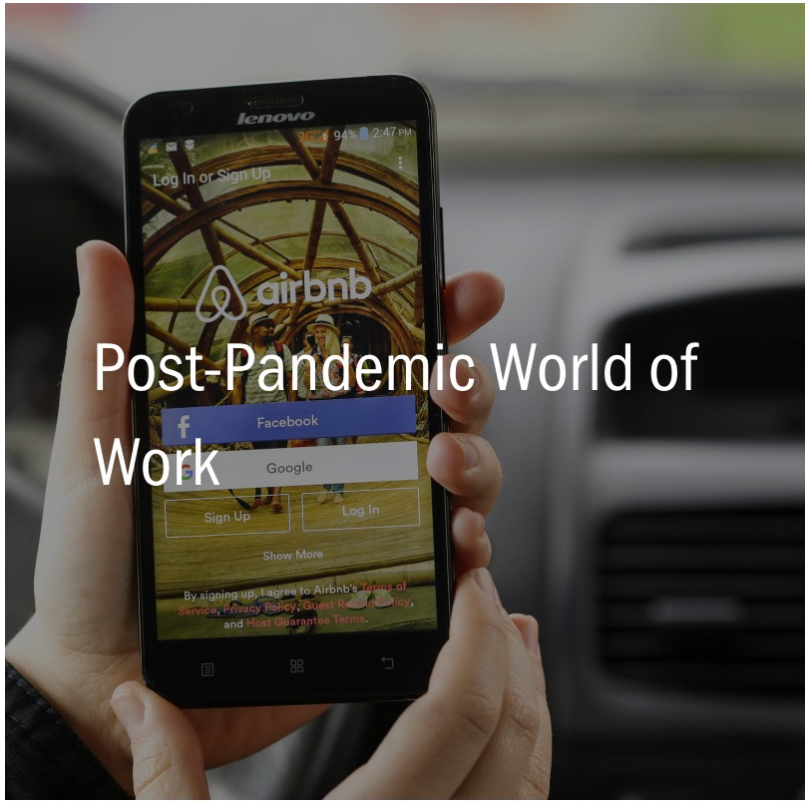
https://biocomplexity.virginia.edu/institute/divisions/social-and-decision-analytics/census



Use Cases - Three Compelling Examples of "Purpose and Use"

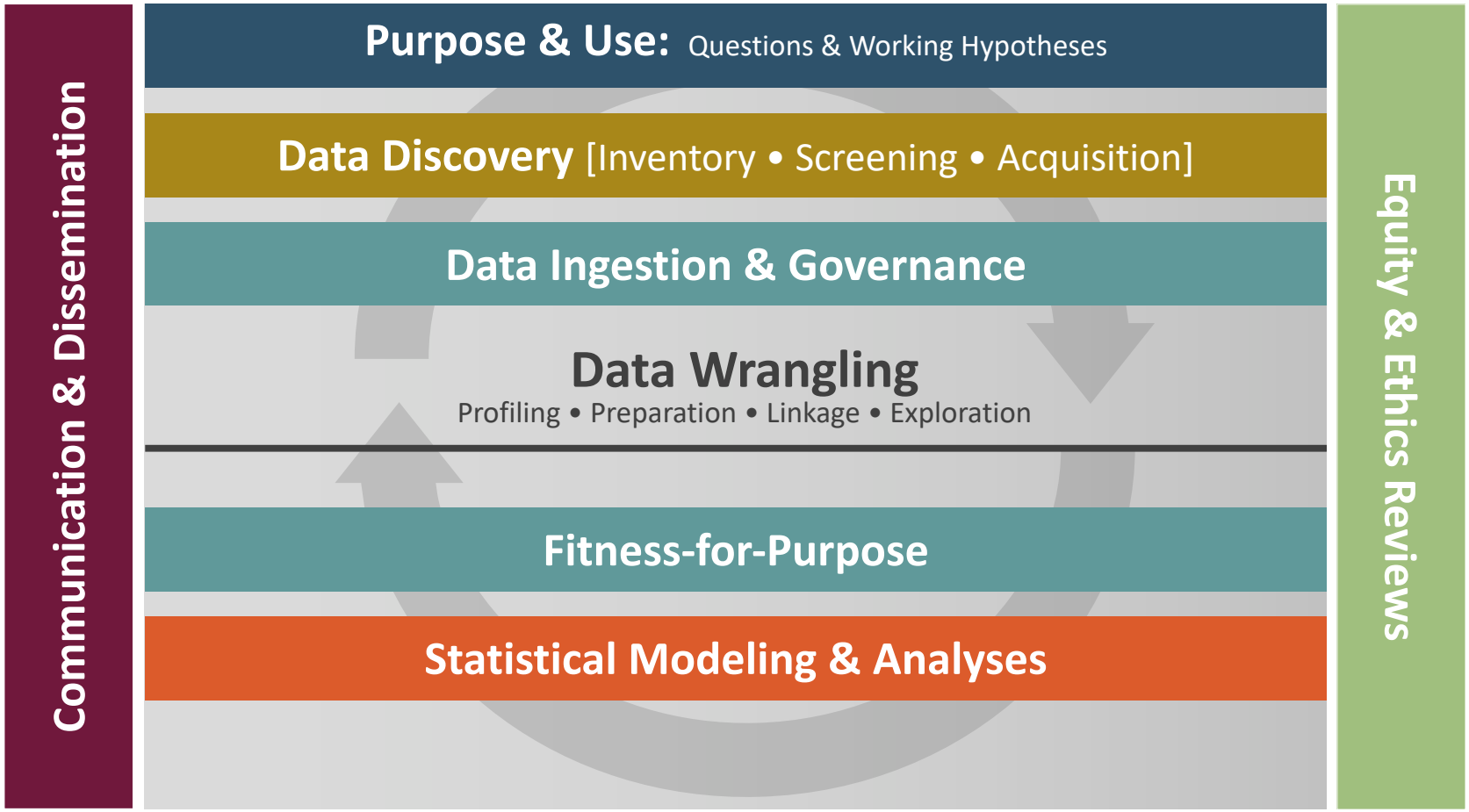
Use Case - a research method to identify, clarify, and organize requirements to create a system or overall system component, such as the Curated Data Enterprise





Framework to Guide Building the CDE Use Case by Use Case

Goal: Illuminate the needed capabilities of the CDE and support its development



Curation applied at every step

Use Case – Vulnerable Populations in Skilled Nursing Facilities

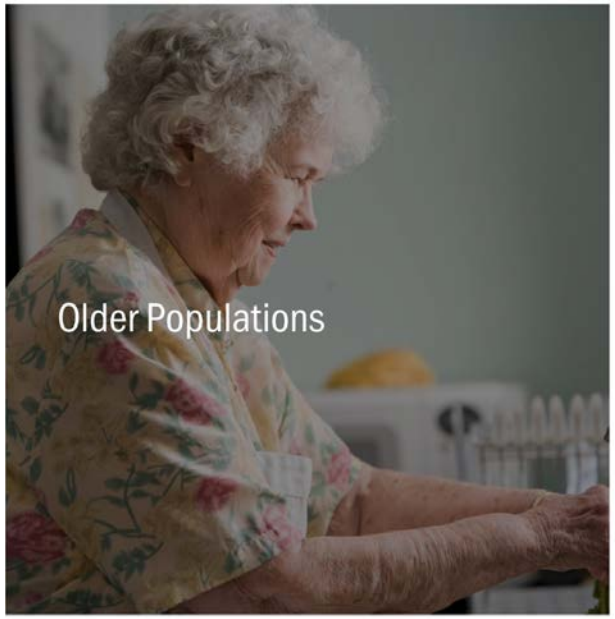
Art of the Possible

Comprehensive picture of care

- **Baseline: Residents, Staff, Owners, Skilled Nursing Facilities (SNF)**
  - Are SNF prepared for emergencies?
  - Can SNF workers get to work?
  - Can communities provide support to SNF during an emergency?

Data Sources

- Flood risks, community asset, travel routes
- CMS administrative data –SNF preparedness
- ACS, CMS, CDC, and NIH surveys



CDE Capabilities to Exercise

- Create risk scores by routes to work (SNF)
- Compute nursing home preparedness
- Identify community assets and resources
- Integrate data with different units of analysis, timing, geography

Conceptual Data Map Aligned to Purpose and Use

Wish List of Variables Constructed from a SNF Literature Search

Census Linked Frames

- Geography
- People
- Jobs
- Businesses

Baseline Curated Data Enterprise

County Level Measures

- Emergency facility services, shelters
- # of nurses per 1000 persons >= 65
- # of home health agencies per 1000 >= 65
- # of hospital beds per 1000 >= 65
- # of empty nursing home beds / # of SNF

Census Tract Level Measures

- **FEMA resiliency**
- CDC SVI
- Climate and Economic Justice Screening Tool
- %<High School
- % disabilities
- % unemployed
- % no vehicle

Skilled Nursing Facility Measures

Residents

Demographic:

- % Minority
- % > age 65
- Mental Health
- Physical Health
- Cognitive Function
- % on Medicaid
- Total # of residents

Workers

Total staff

Hours

- RN hours/ resident day
- LPN hours/ resident day
- CNA hours/ resident day

Turnover:

- Total nursing staff turnover
- # administrators who left

Skilled Nursing Facility

- Deficiencies – emergency & building
- Facility type
- Whether the facility is a chain
- Whether the facility changed hands in the last 12 months

A comprehensive picture of skilled nursing facilities

Facility Safety Assessment
Total # of safety deficiencies and inspections over the last 4 years

Building Knowledge Base (BKB)

- Transportation infrastructure
- How workers get to the nursing home
- Worker demographics
- Where workers live

Severe Climate Event

Data Products

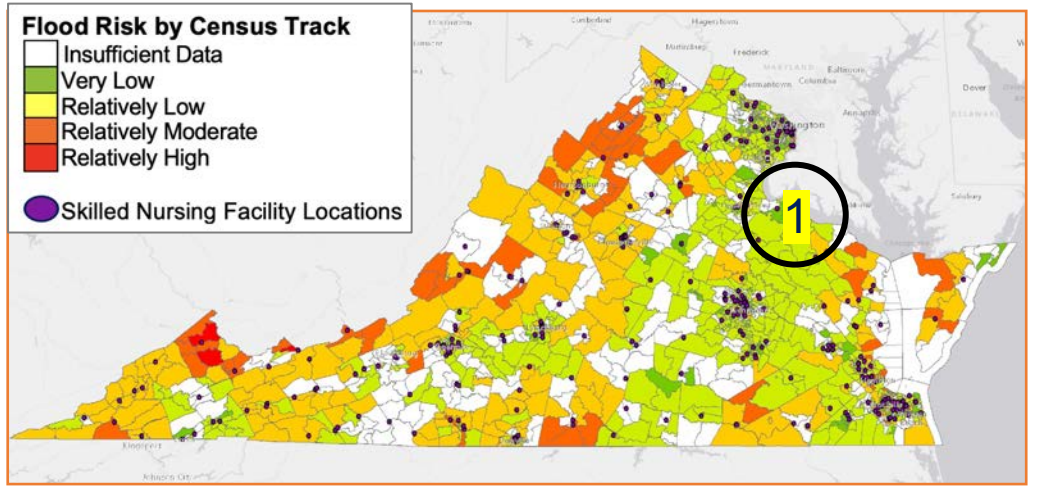
Emergency Preparedness
Total # of emergency preparedness deficiencies and inspections over the last 4 years

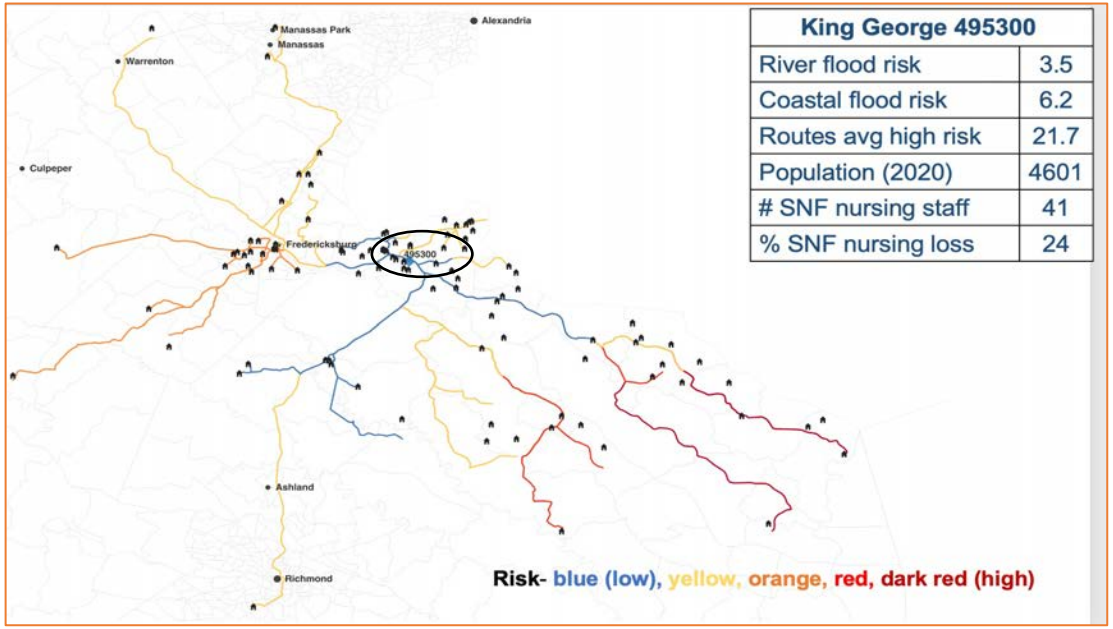
- Coastal and river flooding predictions
- Past hurricane and extreme flooding events

Building on the baseline to address specific questions

Q1: Can SNF workers make it to work during a flooding event?

- **Data discovery –** flooding risks, nursing home locations, road networks
- **Compute** the likely commuting route for each synthetic worker - Identify workers, home locations, nursing home locations, and routes
- **Overlay** the river and coastal flooding by census tracts
- **Estimate flooding risk** along the transportation routes
- **For each SNF, there is a flood risk -** treat that value as a proxy for the probability of workers not getting to work
- **Calculate risk for each SNF** as the percent of workers that cannot make it to work during a flooding event





Q2: Are SNF prepared for emergencies?

CMS Skilled Nursing Home Deficiencies as a proxy for preparedness

Emergency Preparedness (EP) Deficiencies – 4 elements

- Emergency plans
- Communication plan
- Training and testing
- Policies and procedures

Life Safety Code (LSC) Deficiencies - Fire protection requirements

- Designed to provide a reasonable degree of safety from fire, smoke, and panic
- Construction
- Protection
- Operational features

Emergency Preparedness Index

Calculate — Calculate a separate Emergency Preparedness (EP) deficiency index for each Skilled Nursing Facilities

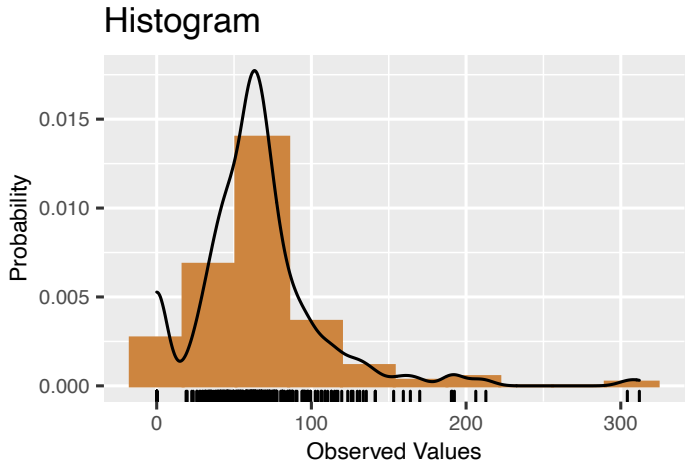
Normalize — 1. Normalize the number deficiencies by the number of SNF inspections over the past four year (2018 – 2022)

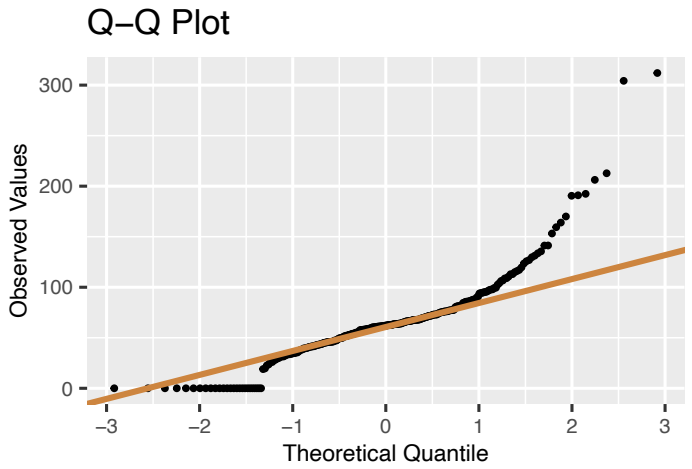
Average — 2. Average the number of days it took to resolve each deficiency

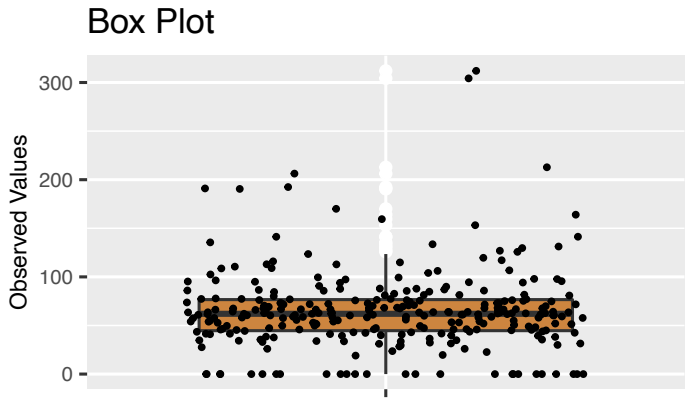
Transform — 3. Transform the letter inspection rating for scope and severity to a number using the *Care Compare Nursing Home 5-Star Quality Rating System,* take average

Add — Add 1 through 3 to get the deficiency index for emergency preparedness (EP) and fire safety (LSC)

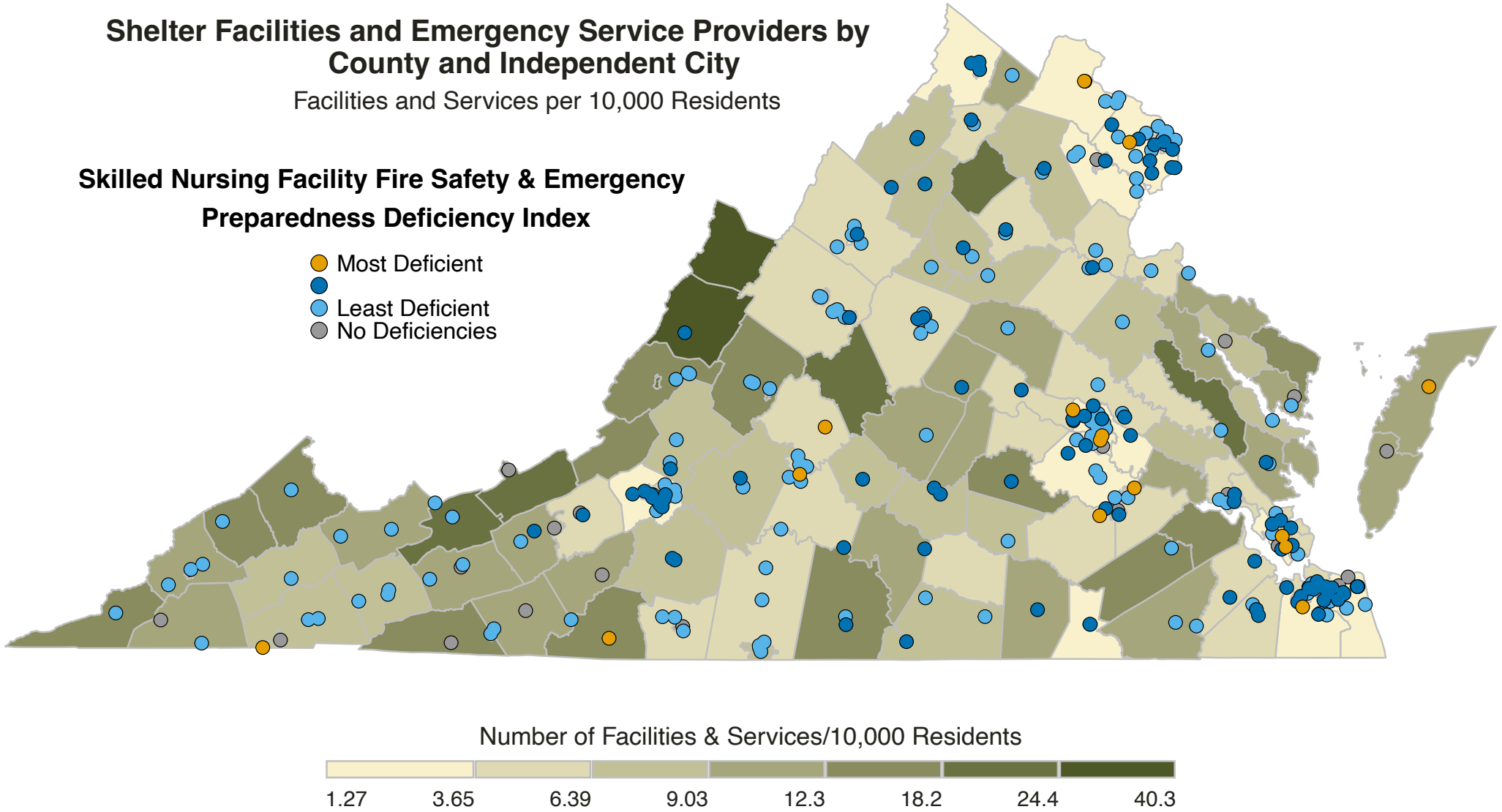
DATA SOURCE: The Centers for Medicare & Medicaid Services Fire Safety Deficiencies
https://data.cms.gov/provider-data/dataset/ifjz-ge4w
Inspection Dates https://data.cms.gov/provider-data/dataset/svdt-c123





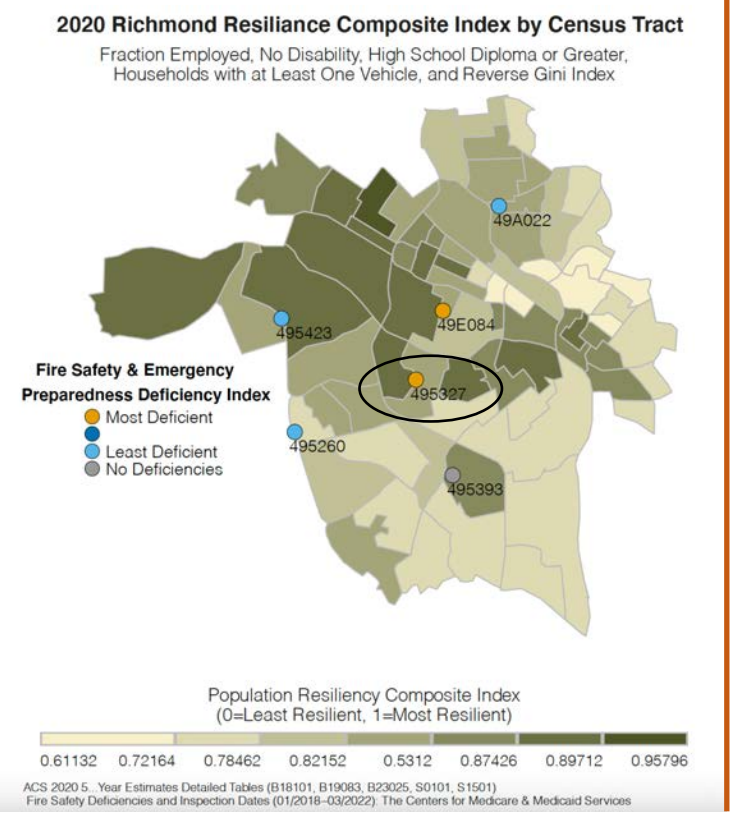


Q3: What is Local Capacity to Mitigate Emergency Events?



Homeland Infrastructure Foundation–Level Data: https://hifld–geoplatform.opendata.arcgis.com/
Fire Safety Deficiencies and Inspection Dates (01/2018–03/2022): The Centers for Medicare & Medicaid Services

If we take a closer look, what can we learn?



At the SNF level, no facility changed **ownership i**n the last year.

- 3 for-profit, 2 not-for-profit, and 1 government owner.

Relatively **high riverine flood risk and low coastal flood risk**.

- On average 18% of the SNF workers will not make it to work if an extreme flooding event occurs.

SNF 495327 had emergency preparedness deficiencies that were wide in scope but not immediate jeopardy; they failed to:

- Conduct tests and exercises for requirements (EP);
- Establish policies and procedures for sheltering (EP); and
- Test and maintain the automatic sprinkler and fire alarm systems (LSC).

City of Richmond SNFs	495260	495327	495393	495423	49A022	49E084
Maximum Riverine Flood Risk	13.80	15.51	14.57	16.39	14.41	14.48
Maximum Coastal Flood Risk	2.91	2.33	1.66	2.28	1.62	2.30
% Nursing Staff Loss if Flood Event	18%	20%	17%	21%	15%	17%
SNF Deficiency Index	Low	High	None	Low	Low	High

GitHub Structure - Skilled Nursing Facilities

Folders, sub-Folders, and sub-sub-Folders

Documents

Literature	Products
Deficiencies	Data Tables**
Emergency Preparedness	Curation Report & Presentations
Nursing Staff Levels	
Ownership	Visualizations
Quality	Processes**
Resilience	

Source Code

Discovery	Analyses

Subfolder: **Discovery**
- File: EDAshape.R
  - Output: eda_emergency_preparedness_deficiency_index.pdf
  - Output: eda_fire_life_safe_code_deficiency_index.pdf
- File: EDA_flatviolin.R
  - Output: eda_snf_deficiency_index_by_hazard.pdf
- File: EDA_boxplots.R
  - Output: county_population_indicator_resilience_box_plots.pdf
- File: EDA_Choropleths_Resilience_Indicator_Variables.R
  - Output: census_tract_population_16_and_over_unemployed_choropleth.pdf
  - Output: census_tract_population_no_hs_diploma_choropleth.p
  - Output: census_tract_population_with_disability_choropleth.pdf
  - Output: county_old_age_dependency_choropleth.pdf

Subfolder: **Analyses**
- File: SNF_Deficiency_Index.R
  - Output: va_snf_deficiency_indices_k_e.csv
- File: VA_Population_Resilience_Index_Census_Tract.R
  - Output: census_tract_population_resilience_index_choropleth.pdf
  - Output: va_census_tract_population_resilience.csv
- File: VA_Population_Resilience_Index_County.R
  - Output: county_population_resilience_index_with_gini_choropleth.pdf
  - Output: county_population_indicator_resilience_box_plots.pdf
  - Output: va_county_population_resilience.csv
- File: VA_Probability_of_Getting_to_SNF.R
  - Output: va_snf_estimated_average_daily_nursing_staff_during_extreme_flood_event.csv
  - Output: snf_estimate_daily_nursing_staff_during_climate_event.pdf

Data

SNF***	Community
Facility*	Demographic
Nursing Staff*	Climate Change Risk*
Residents*	Resilience*
Owners*	

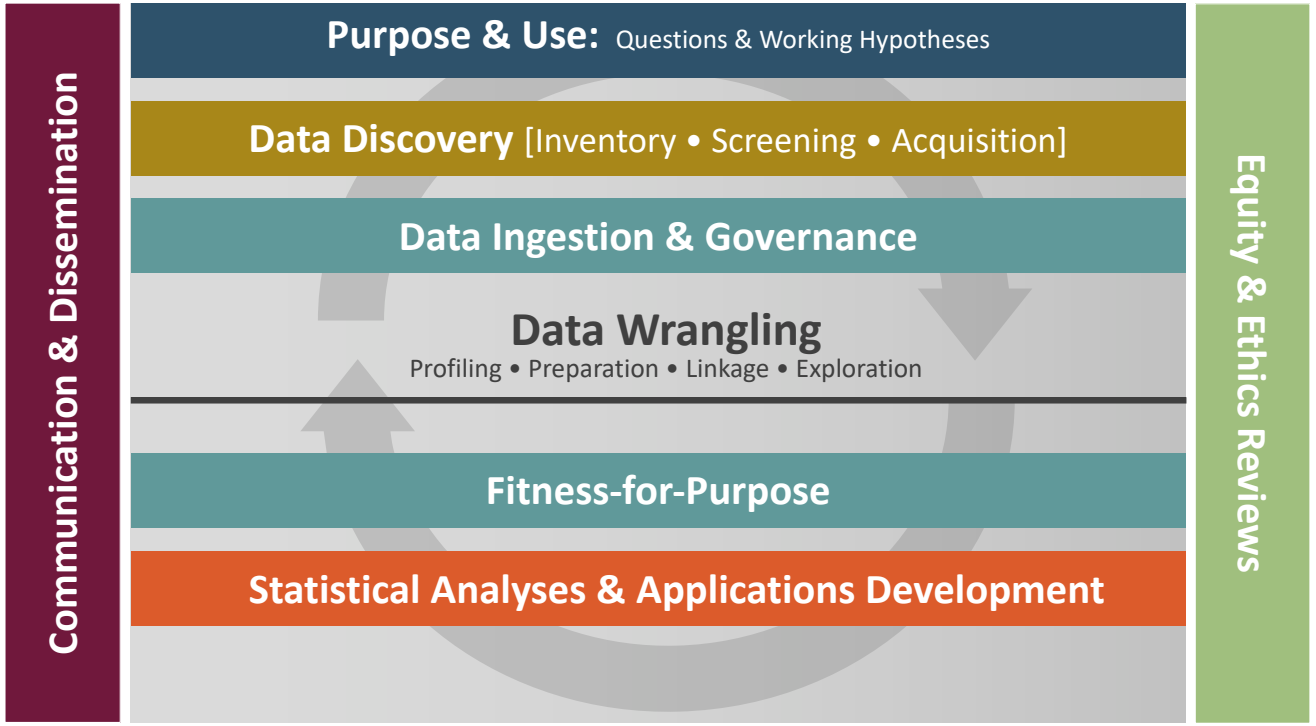
https://github.com/uva-bi-sdad/census_cde_demo_2/

*Each folder contains a meta data folder with data dictionaries and technical documents.
** Contains additional subfolders. ***SNF is Skilled Nursing Facilities

Use Case Research Selection Criteria

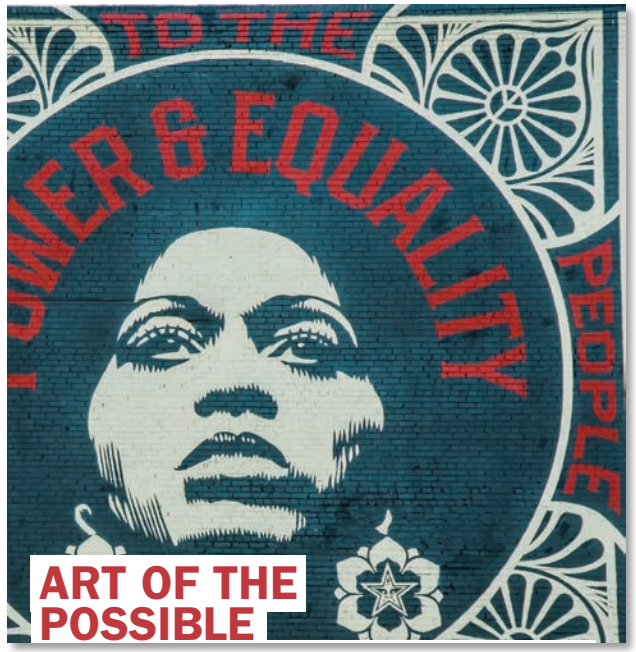
Building the CDE Use Case by Use Case
What are criteria to select Use Cases to support CDE development?
How could a Use Case exercise some aspect of the CDE?

- Relevancy
- Multiple data sources
- Computing requirements
- Subject Matter Experts
- Data Partnerships
- Ethics and equity
- Curation



Use and Reuse of CDE Components

Resolve once, use many times



Use cases inform CDE component development and reusability

- Data Sources
- Functions
- Applications

Data sources

- Synthetic population estimates
- SNF location and characteristics
- Baseline community emergency preparedness assets

Functions

- Creating transportation routes
- Computing the probability of flood risks
- Estimating SNF resiliency & community resilience

Applications

- Curation - Organize and document data to enable the discovery and reuse in research and data production
- Wrangle & integrate many types and sources of data
- Understand differences in similar measures

Seizing Opportunities

*"**As we advance into the 21st century**, we are experiencing increased demand for our data, struggling with challenges to traditional data collection methods, and **exploring rich new data sources 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."*** *(Keller et al. 2022)*

Ron Jarmin, Census Deputy Director, and Chief Operating Officer

Acknowledgements and Thanks!

UVA CDE Working Group

- Ken Prewitt
- Sallie Keller
- Steve Jost
- John Thompson
- Sarah Nusser
- Joe Salvo
- Matthew Snipp
- Zhengyuan Zhu
- Chris Barrett
- Stephanie Shipp

UVA Census Demo Team

- Vicki Lancaster
- Aaron Schroeder
- Samarth Swarup
- Henning Monteith
- Dawn Xia
- Stephanie Shipp

Thanks to Census Bureau advisors and collaborators

Selected References

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

Keller S, Shipp S. **A New Approach for the 2030 Decennial Census - A 21st Century Census for a 21st Century Society**, Technical Report TR 2022-022. Proceedings of the Biocomplexity Institute, University of Virginia; 2021 February. https://doi.org/10.18130/3was-x496

Keller S, Shipp S. **Distribution of Census-Guided Federal Funds - Establishing the Feasibility of an Essential-Data Decennial Census and 21st Century Census Information Platform**, Technical Report BI-2022-024. Proceedings of the Biocomplexity Institute, University of Virginia; 2022.https://doi.org/10.18130/xavc-v02

Keller S, Shipp S. **Redistricting Congressional and State Legislative Districts - A Proposal for an Essential-Data Count of the Resident Population for the Decennial Census and 21st Century Census Information Platform**. Technical Report TR 2022-023. Proceedings of the Biocomplexity Institute, University of Virginia. 2021. https://doi.org/10.18130/qyme-ty93

Pristavec T, Shipp S, Keller S. **Using Administrative Records in Official Statistics and Social Science Research,** Technical Report BI 2022-025. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. https://doi.org/10.18130/b7b9-x755

Salvo J, Shipp S, Zhang S. **Building a Case Study of Domestic Migration and the Curated Data Enterprise** 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

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

Shipp S, Salvo J, Zhang S. **Understanding the State of Nursing Homes and Residential Care Facilities - A Use Case to Support the Development of the Curated Data Enterprise**. Technical Report 2022-028. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. https://doi.org/10.18130/5d6x-f676

A 21st Century Census Curated Data Enterprise

Building the CDE Use Case by Use Case

THOUGHTS?