
Evolving a Data Enterprise to Support Relevant, Timely, and Equitable Statistical Products— *We Need Your Help!*

Sallie Ann Keller, Ph.D., Associate Director and Chief Scientist

Research and Methodology Directorate

U.S. Census Bureau

May 16, 2023

*The views expressed in this perspective are those
of the presenter and not the Census Bureau.*

Forces Driving Innovation Today

“
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.**”

Ron S Jarmin, PhD., Deputy Director
U.S. Census Bureau



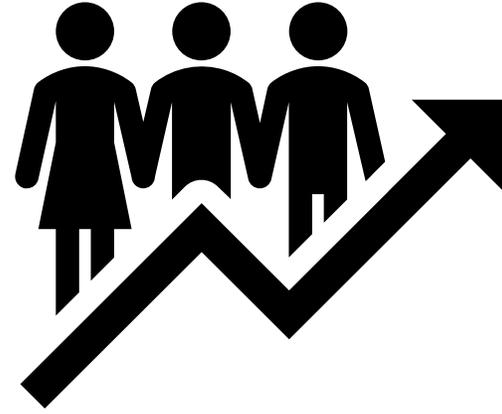
Flipping the Focus

Determine what information stakeholders need to reach their objectives

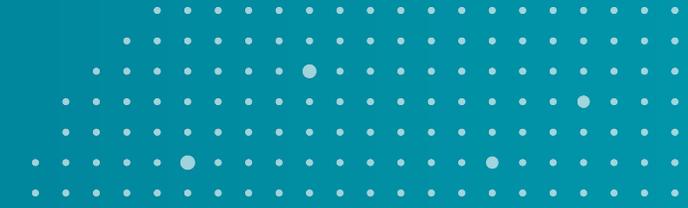
From there, shape the statistical products to be developed



Statistical
Product First
Approach



Statistical Product First Approach



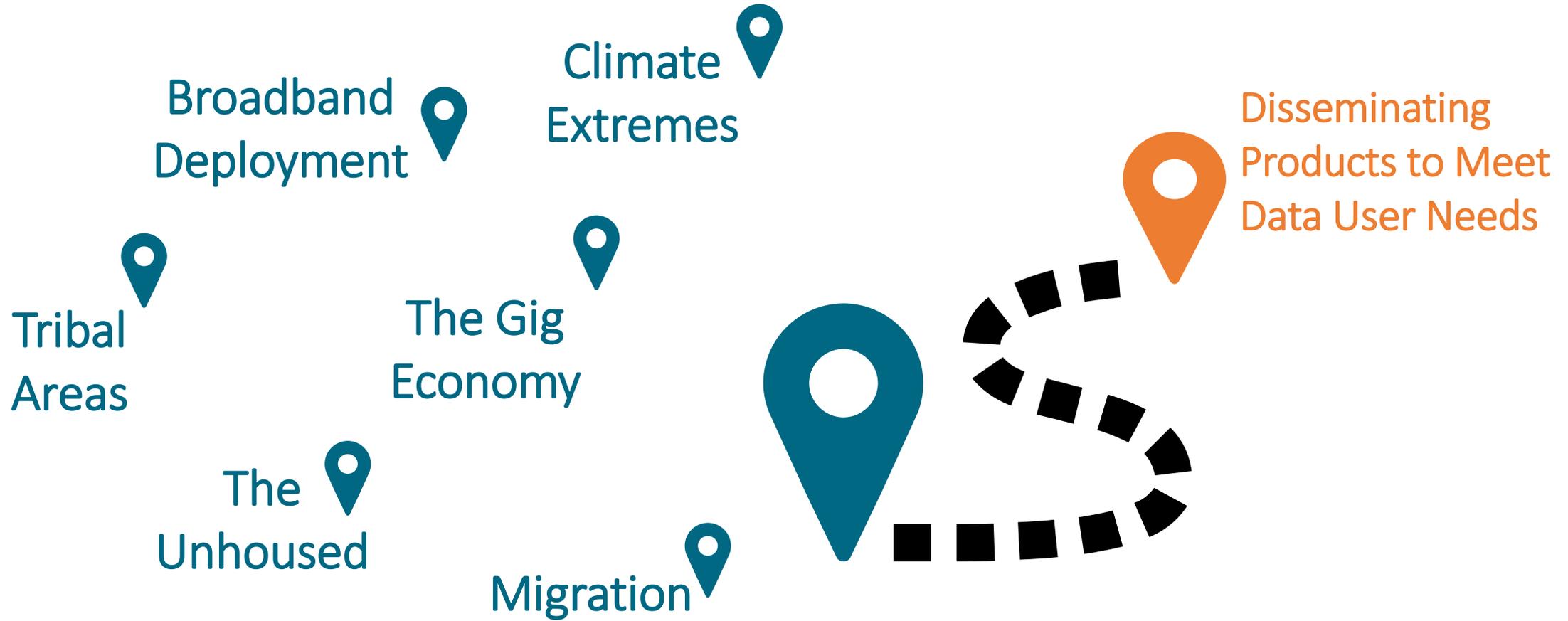
Ensures Data Support Purposes and Uses

Purposes
and Uses



Disseminating
Products to Meet
Data User Needs

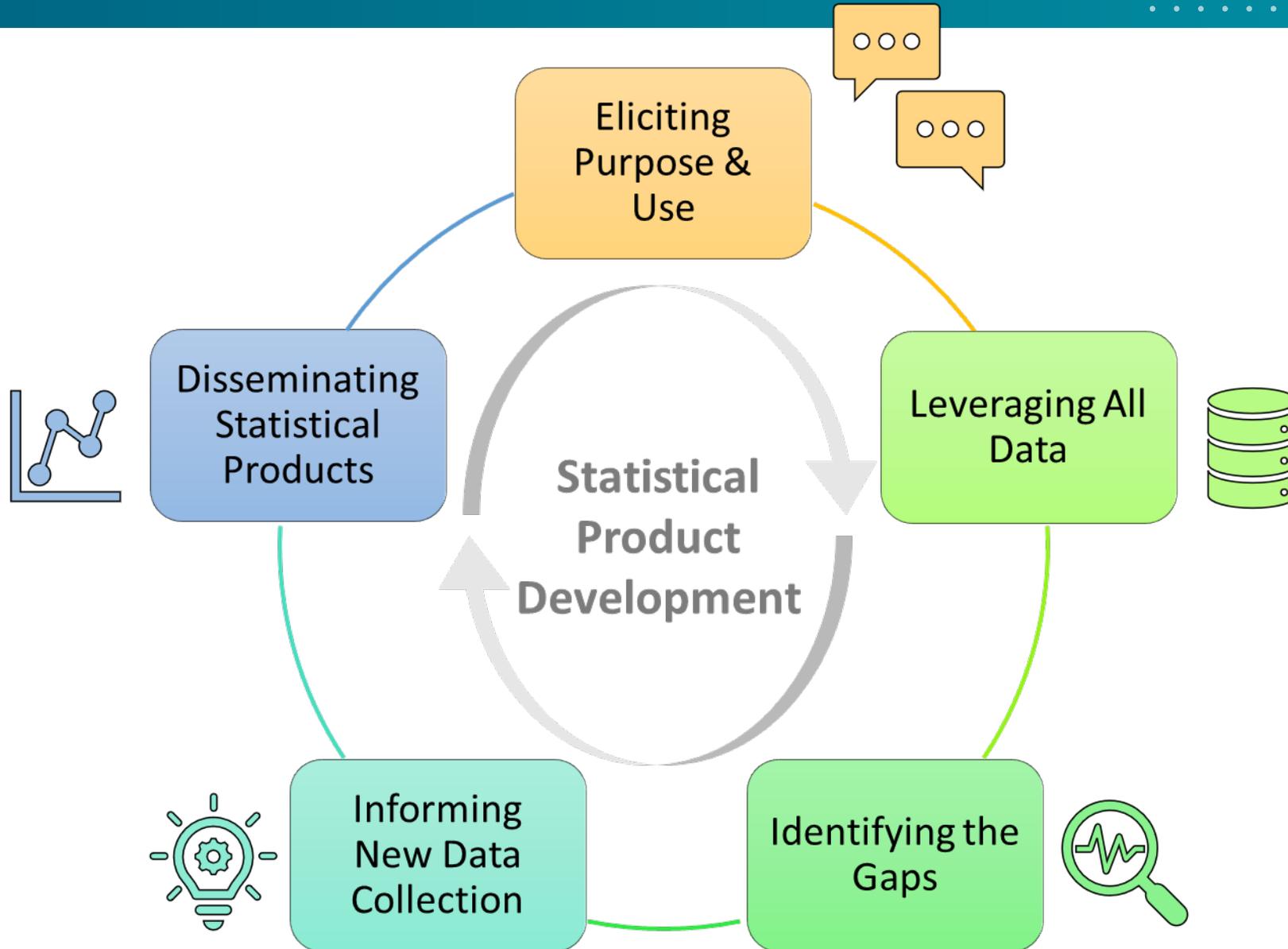
What are some purposes and uses?



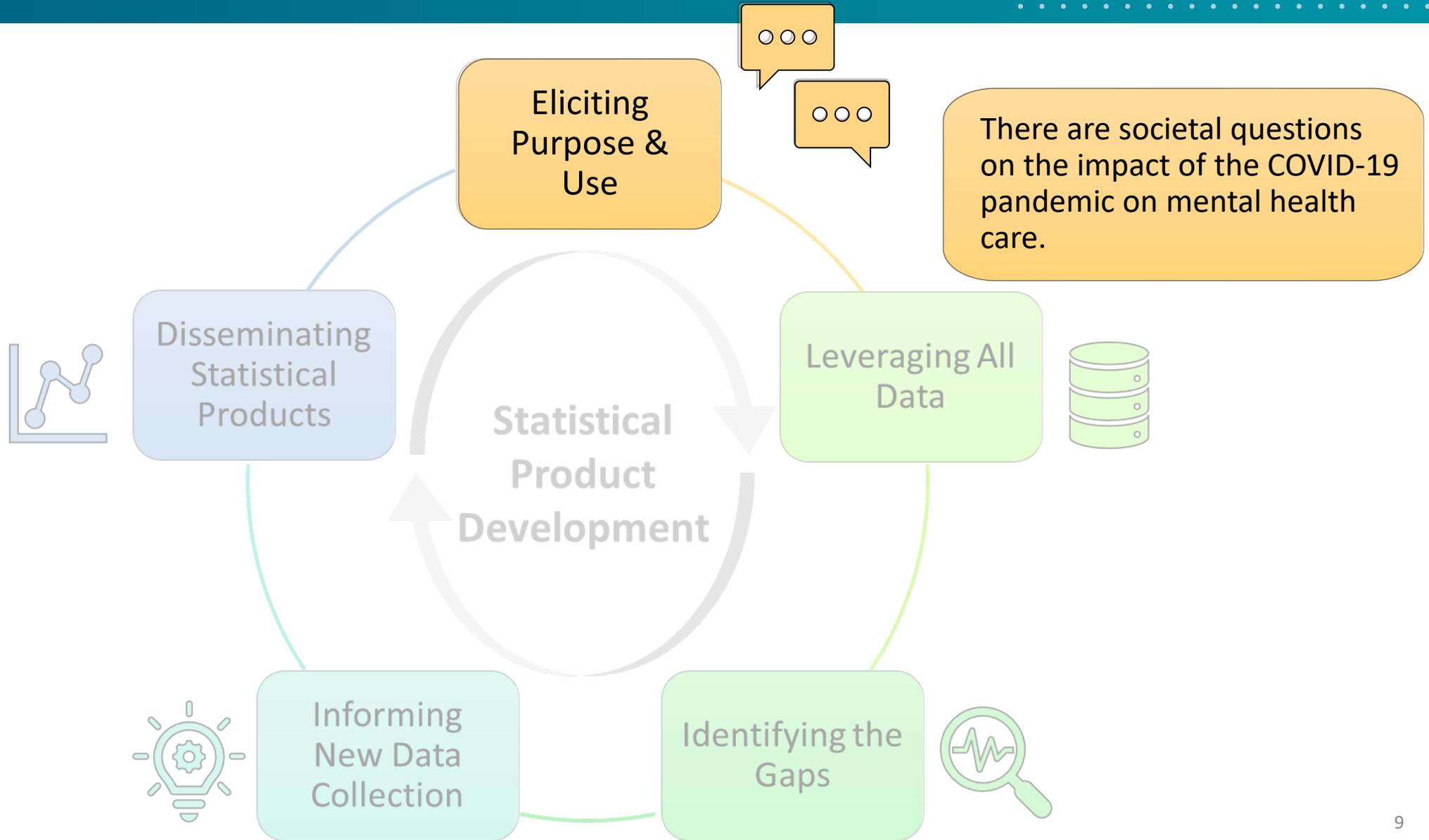
What are some purposes and uses?



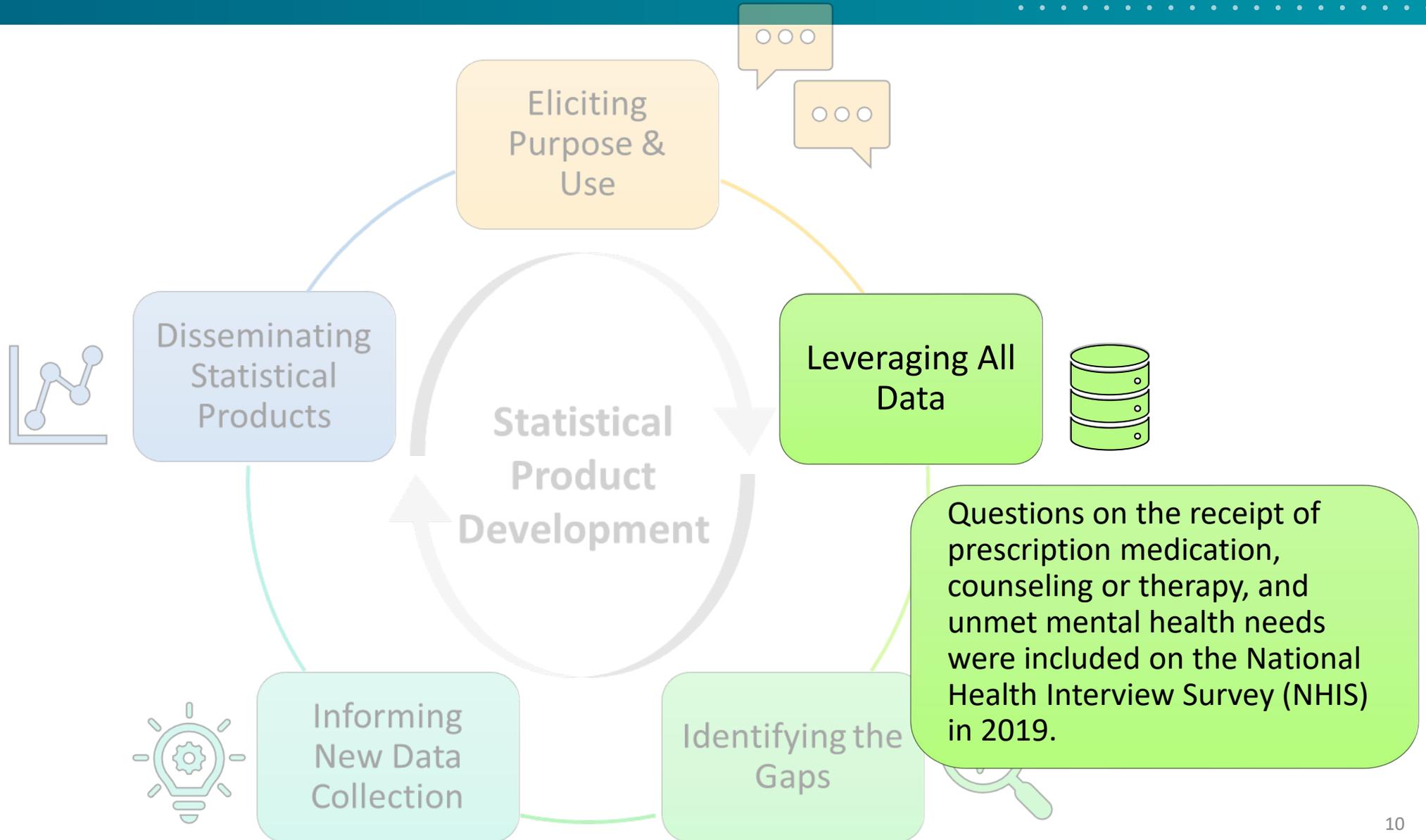
The Continuous Journey Ahead



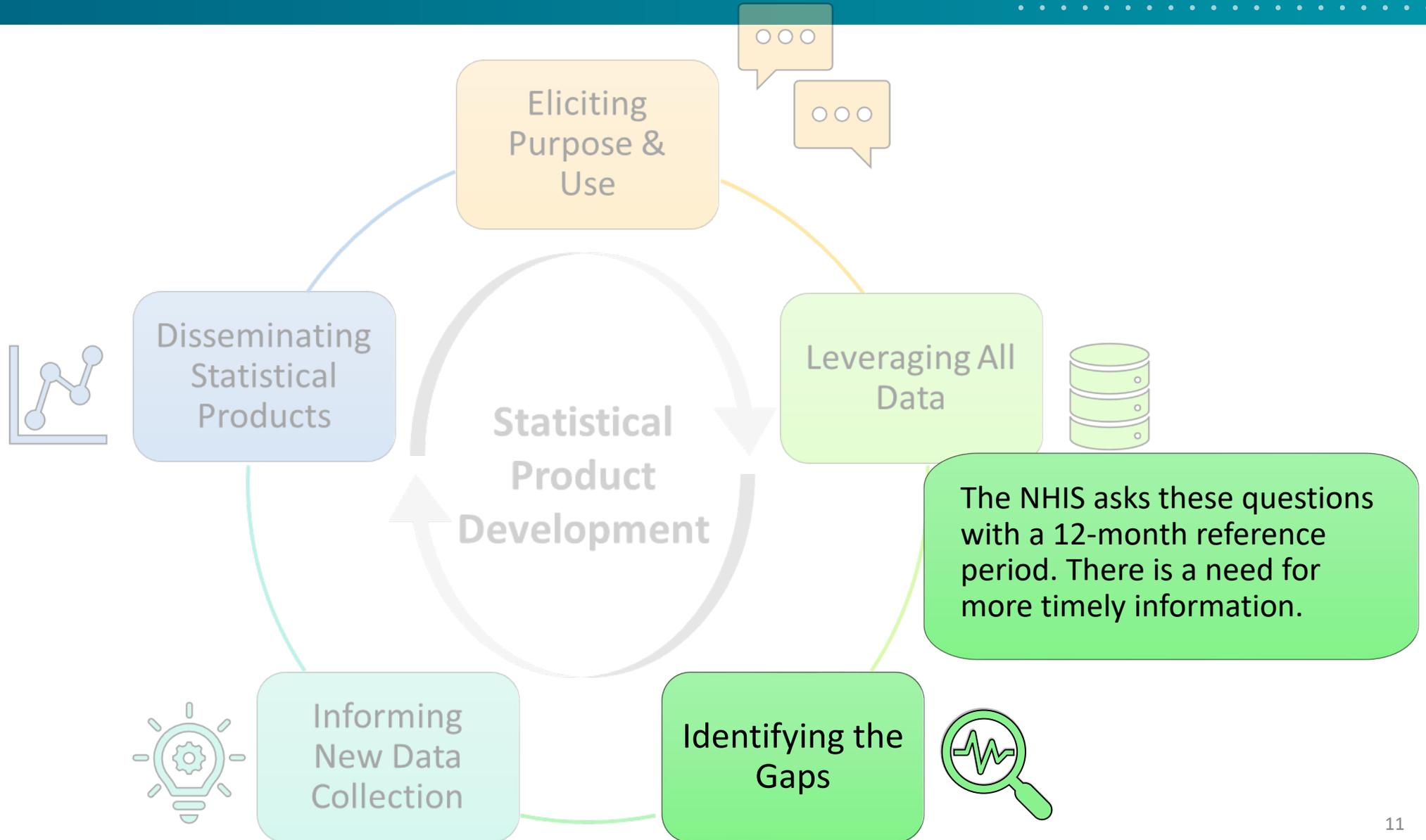
Exemplar: Measurement of Mental Health



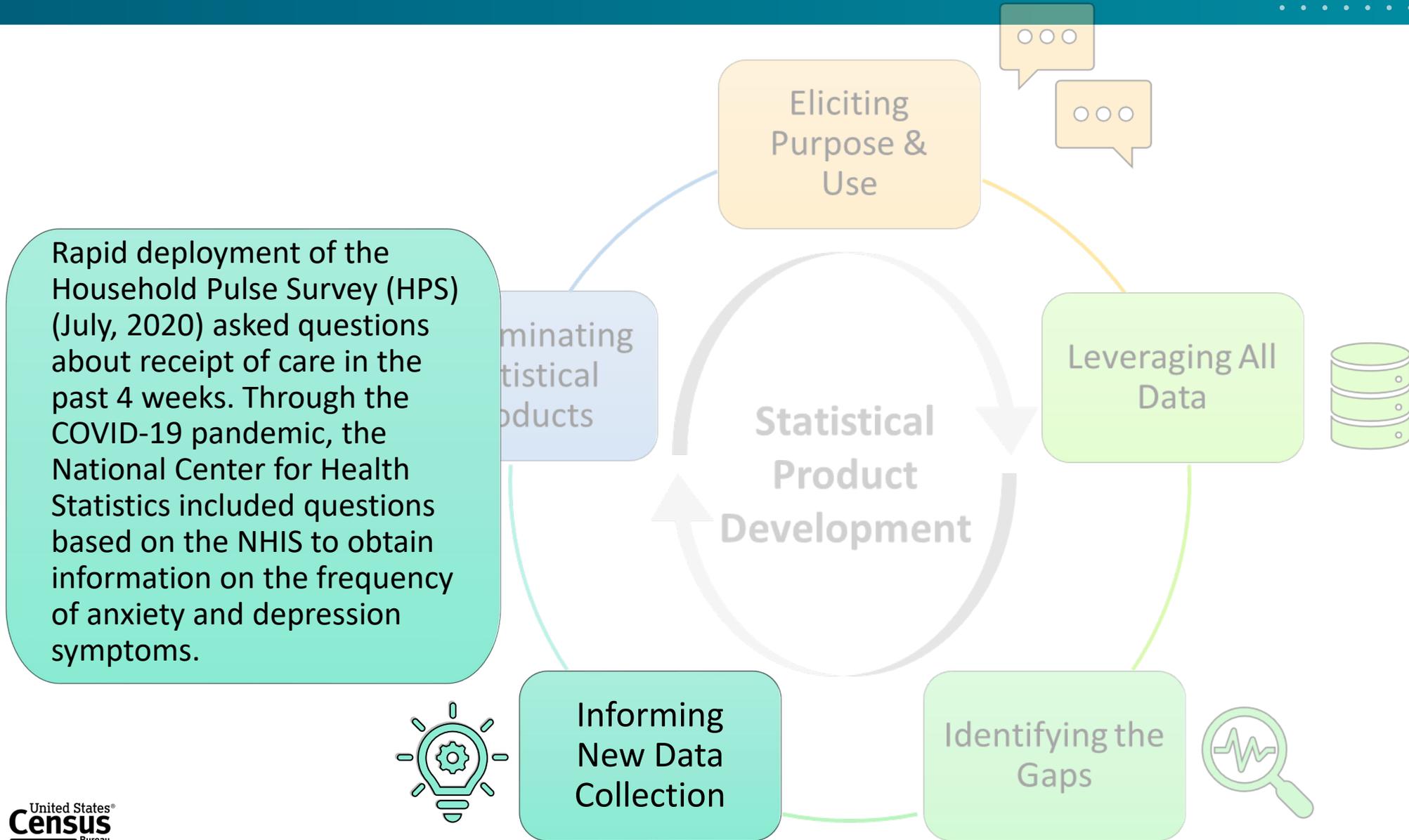
Exemplar: Measurement of Mental Health



Exemplar: Measurement of Mental Health

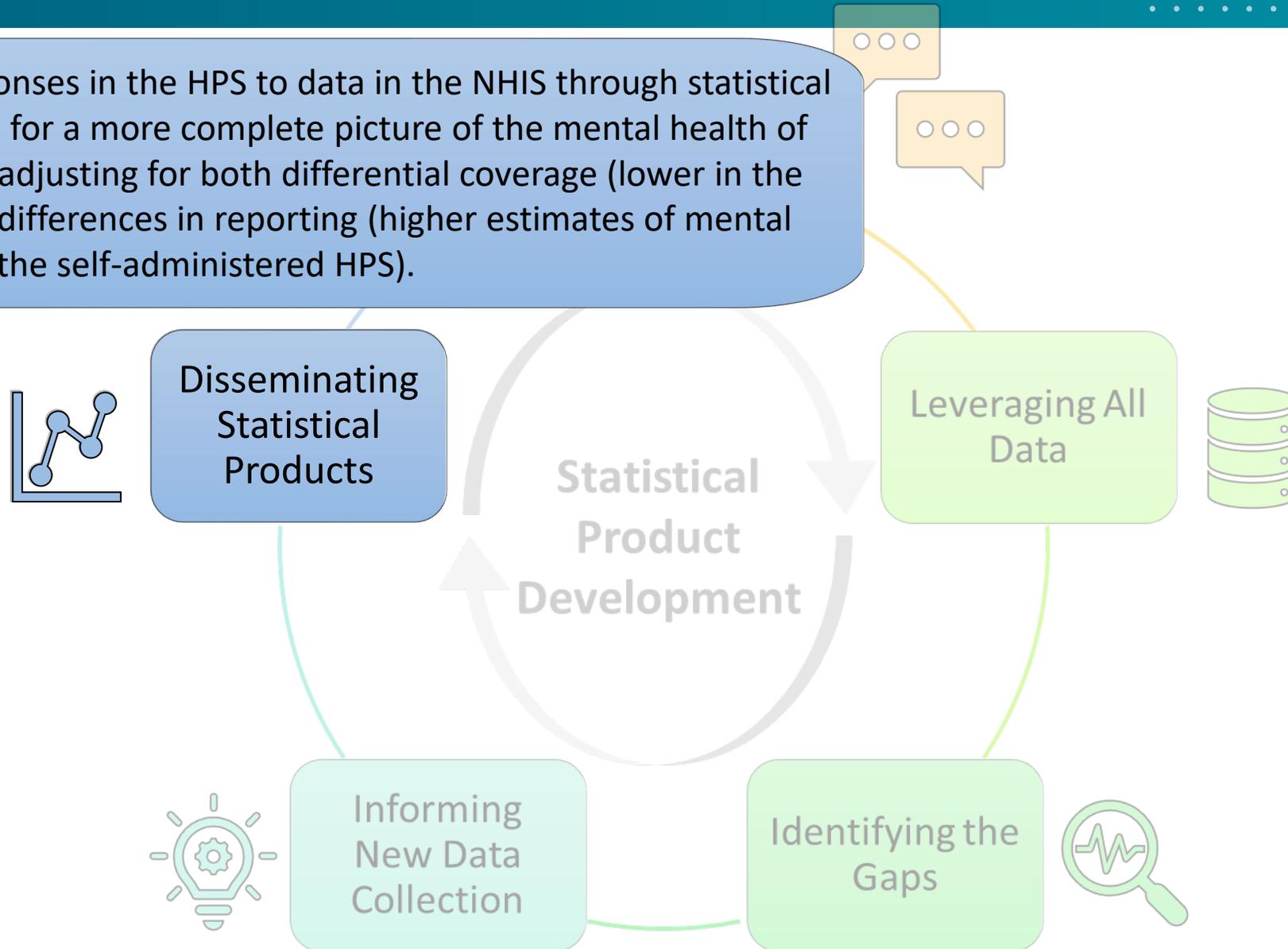


Exemplar: Measurement of Mental Health

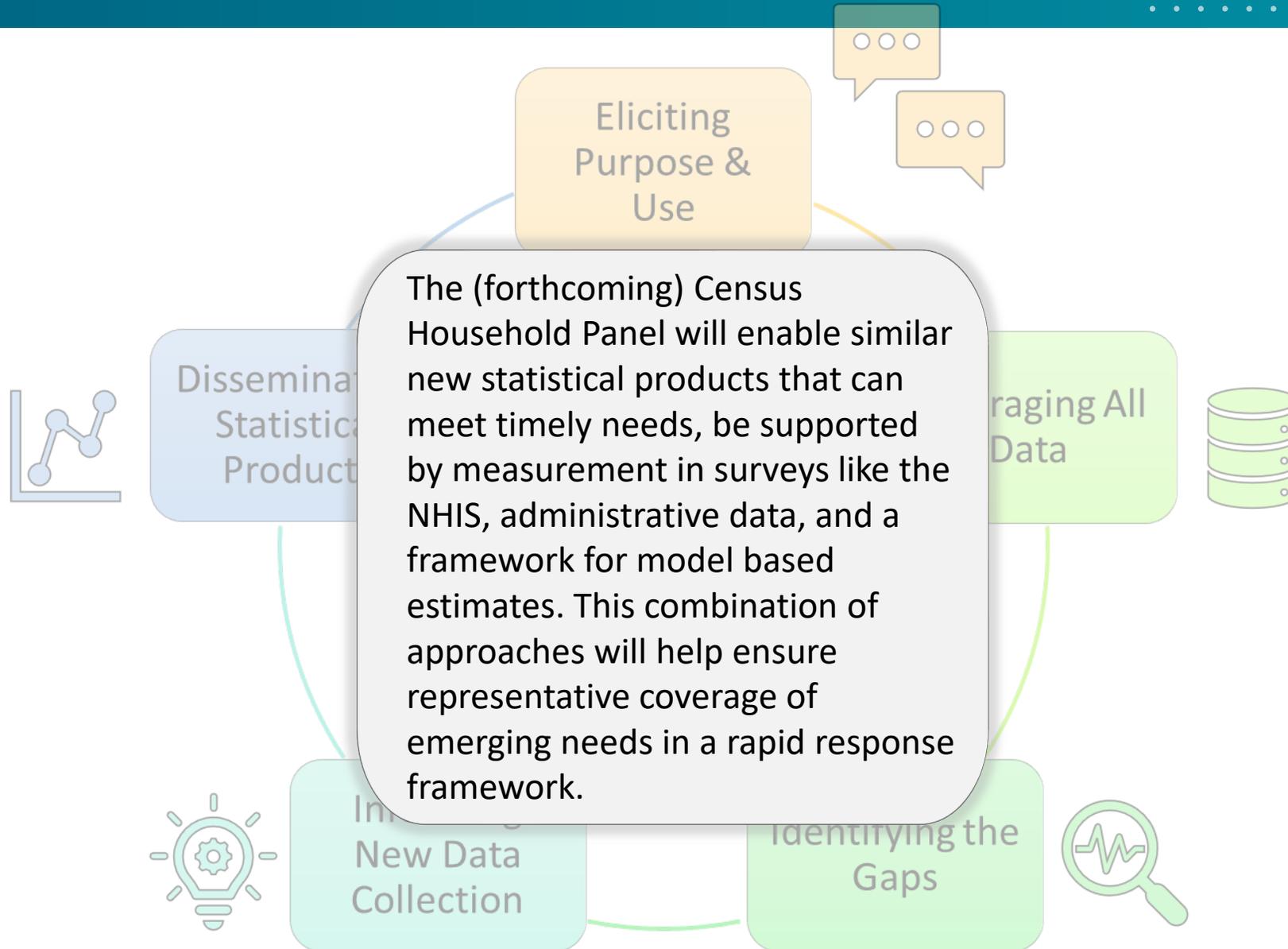


Exemplar: Measurement of Mental Health

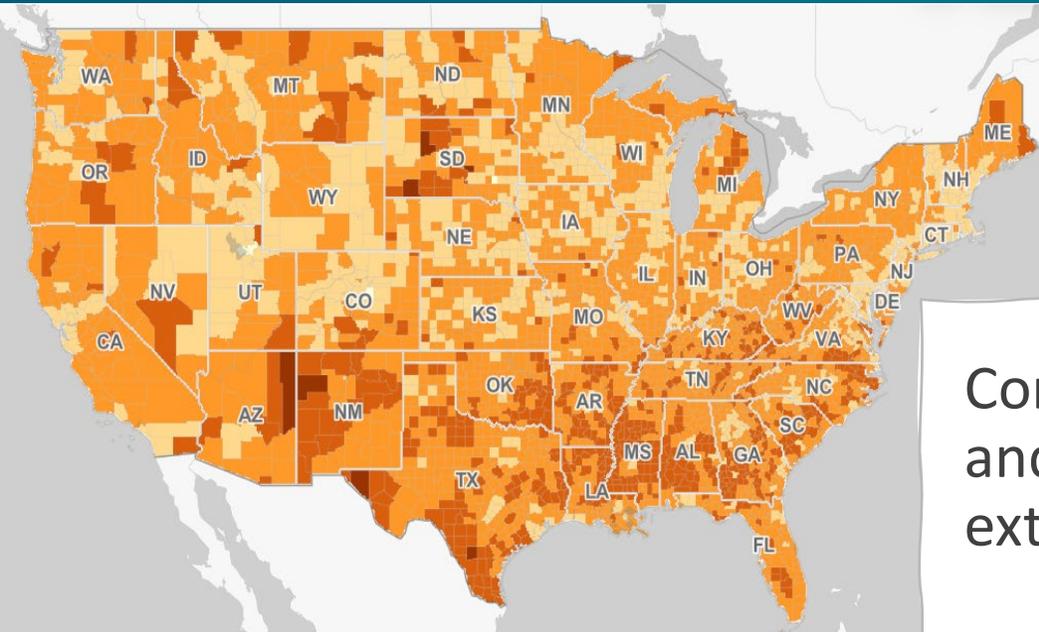
Calibrating responses in the HPS to data in the NHIS through statistical modeling allows for a more complete picture of the mental health of the population, adjusting for both differential coverage (lower in the HPS) and mode differences in reporting (higher estimates of mental health issues in the self-administered HPS).



Exemplar: Measurement of Mental Health



Exemplar: Community Resilience Estimates (CRE) for Heat



Eliciting
Purpose &
Use

Community resilience is the capacity of individuals and households within a community to absorb the external stresses of a disaster.

The CRE was published as an experimental data product to provide information about the COVID-19 pandemic but also to garner feedback from data users on the quality and usefulness of the new product.

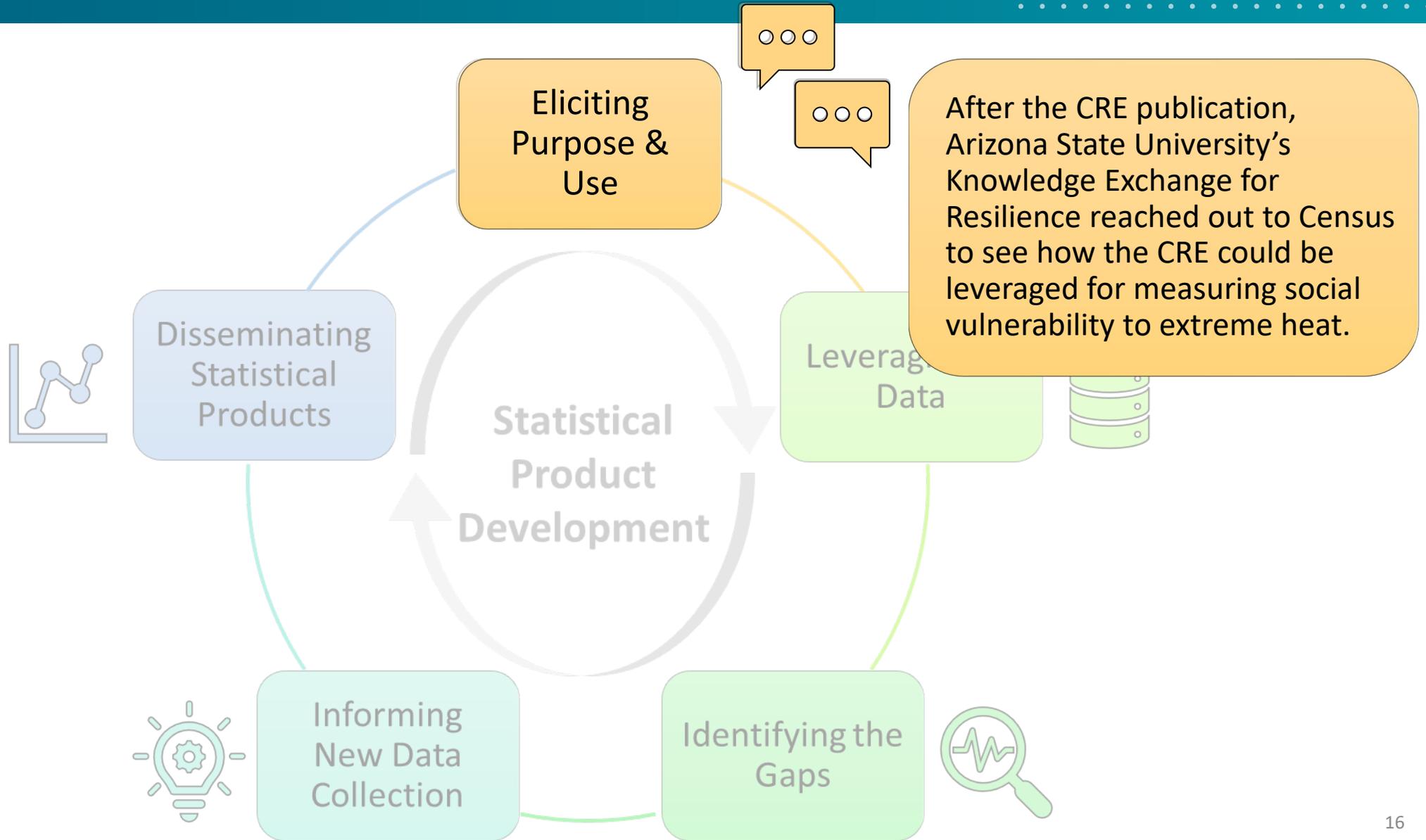


Informing
New Data
Collection

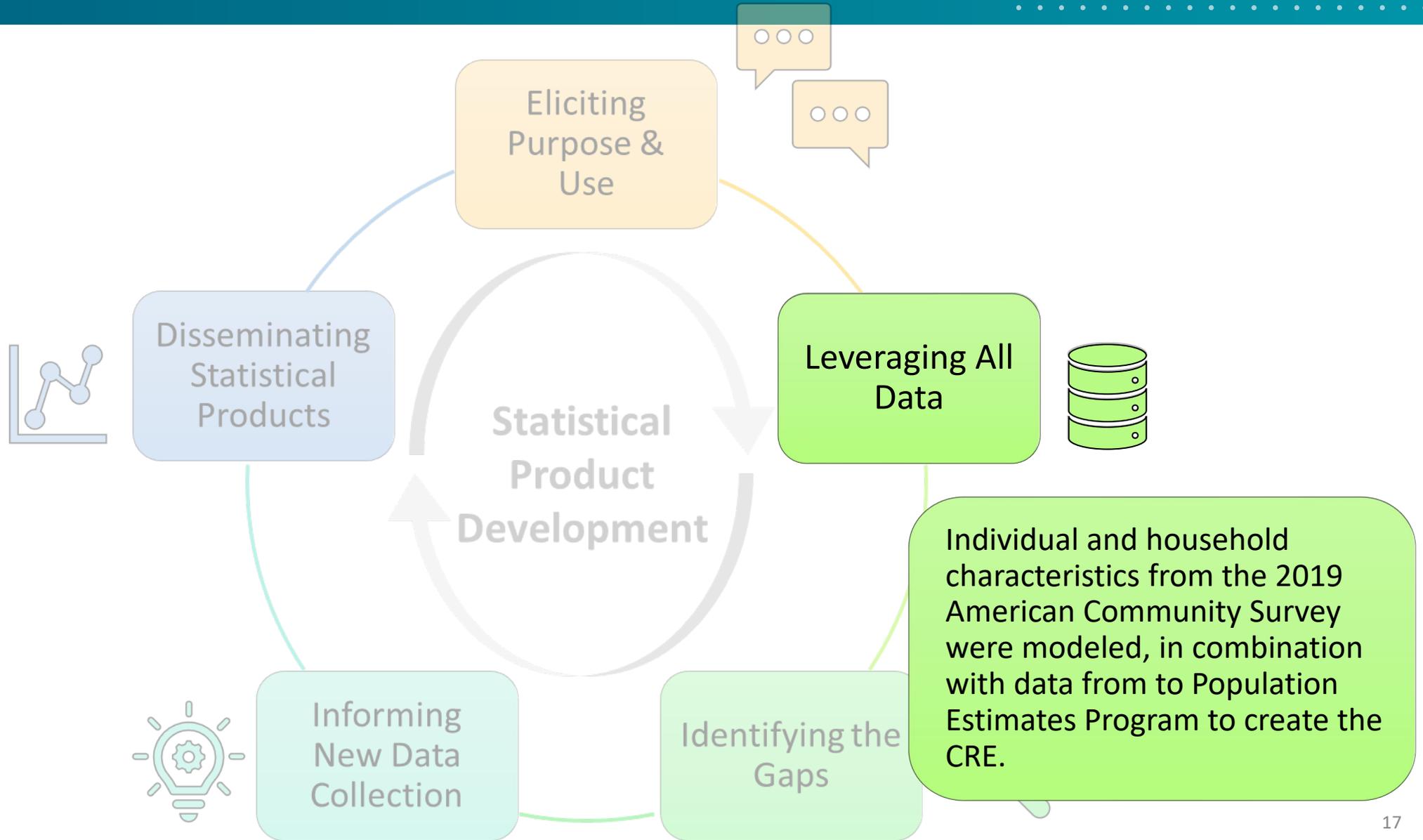
Identifying the
Gaps



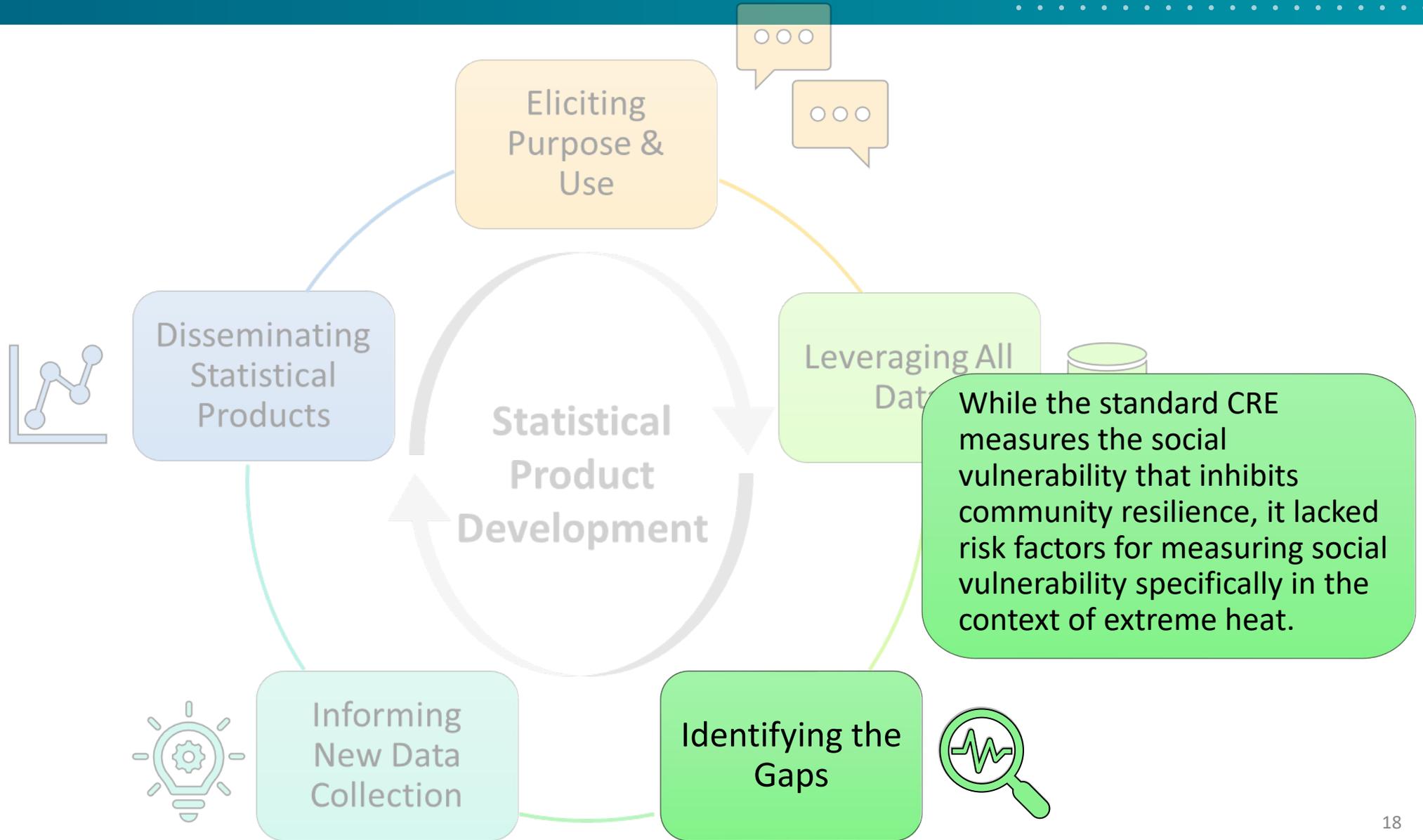
Exemplar: CRE for Heat



Exemplar: CRE for Heat



Exemplar: CRE for Heat



Exemplar: CRE for Heat

With consultation from Arizona State University's Knowledge Exchange for Resilience, the CRE was updated for vulnerability to extreme heat.

Three of the ten standard CRE risk indicators were modified, taking into account:

- housing structure type,
- commute type, and
- whether the household's housing costs were greater than 50%.



Informing New Data Collection

Eliminating Statistical Products

Statistical Product Development

Eliciting Purpose & Use



Leveraging All Data

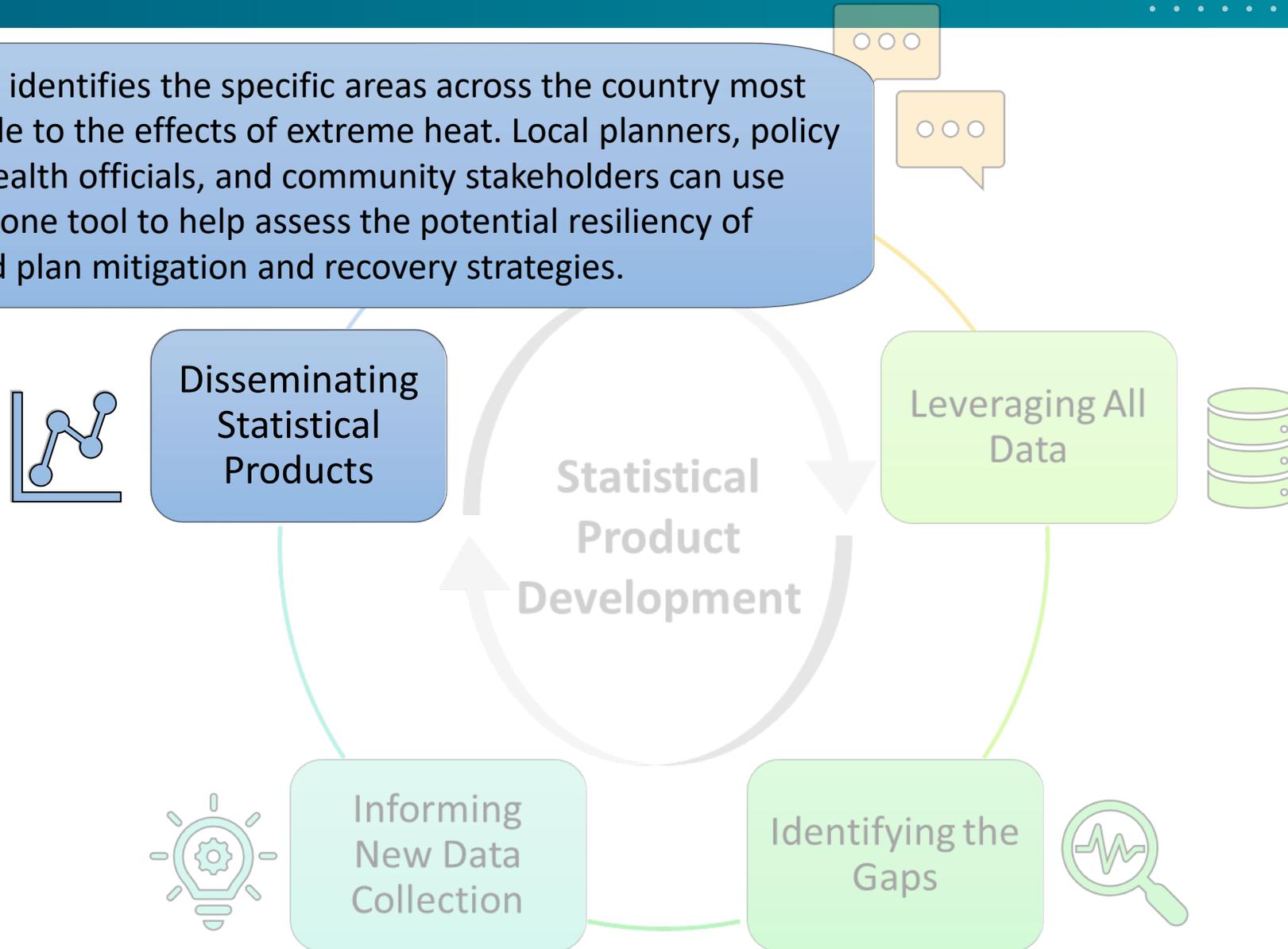


Identifying the Gaps

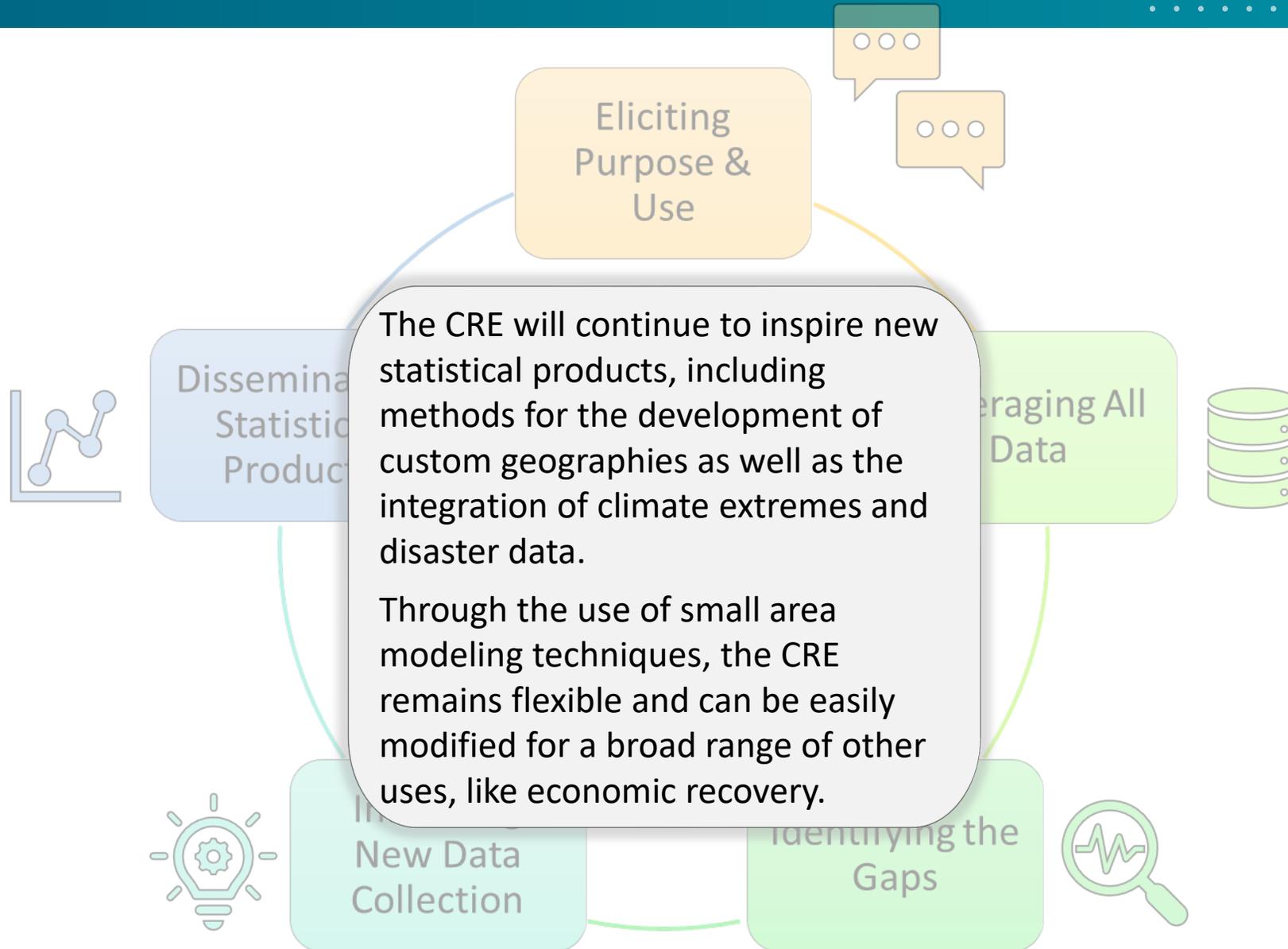


Exemplar: CRE for Heat

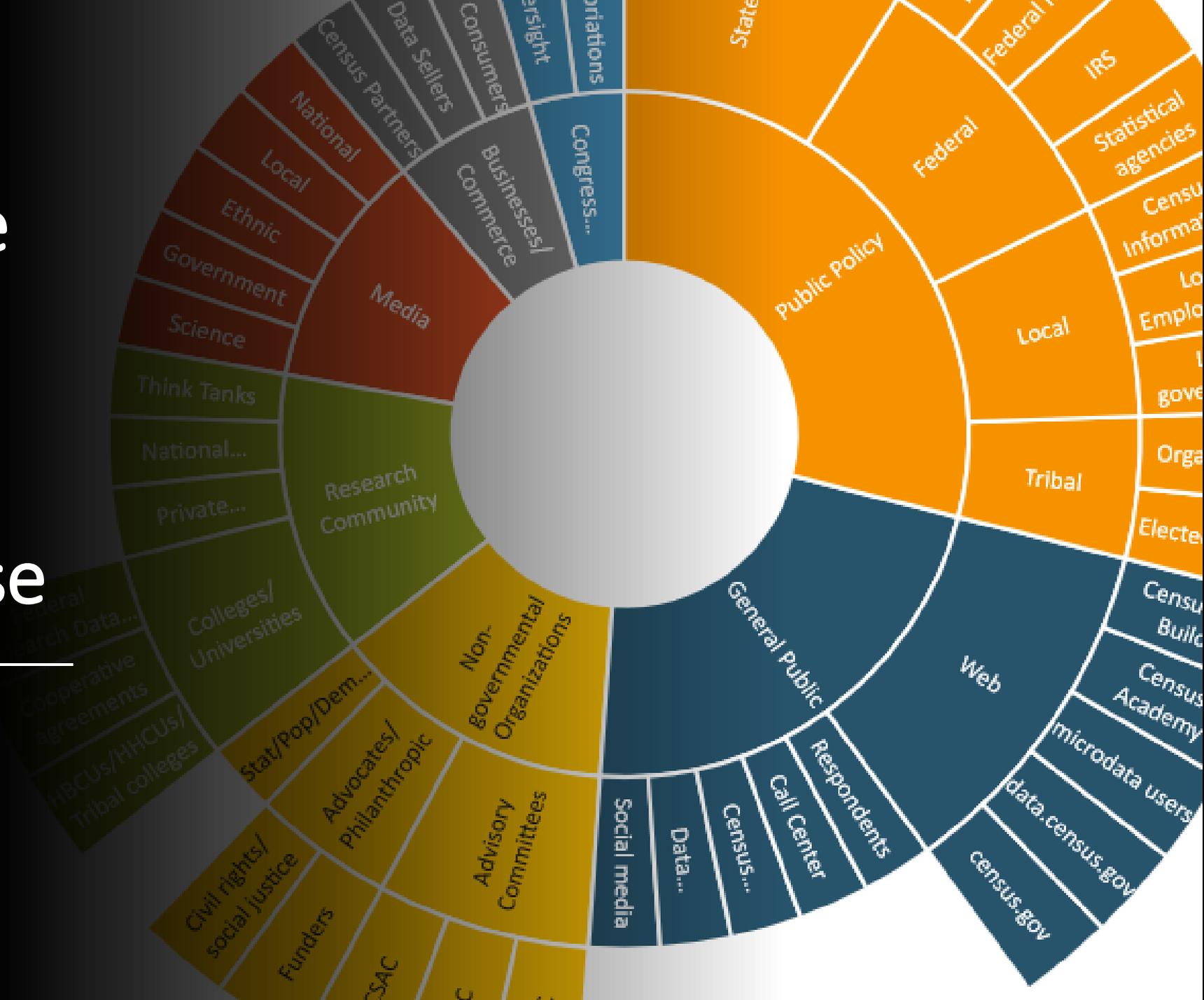
The CRE for Heat identifies the specific areas across the country most socially vulnerable to the effects of extreme heat. Local planners, policy makers, public health officials, and community stakeholders can use the estimates as one tool to help assess the potential resiliency of communities and plan mitigation and recovery strategies.



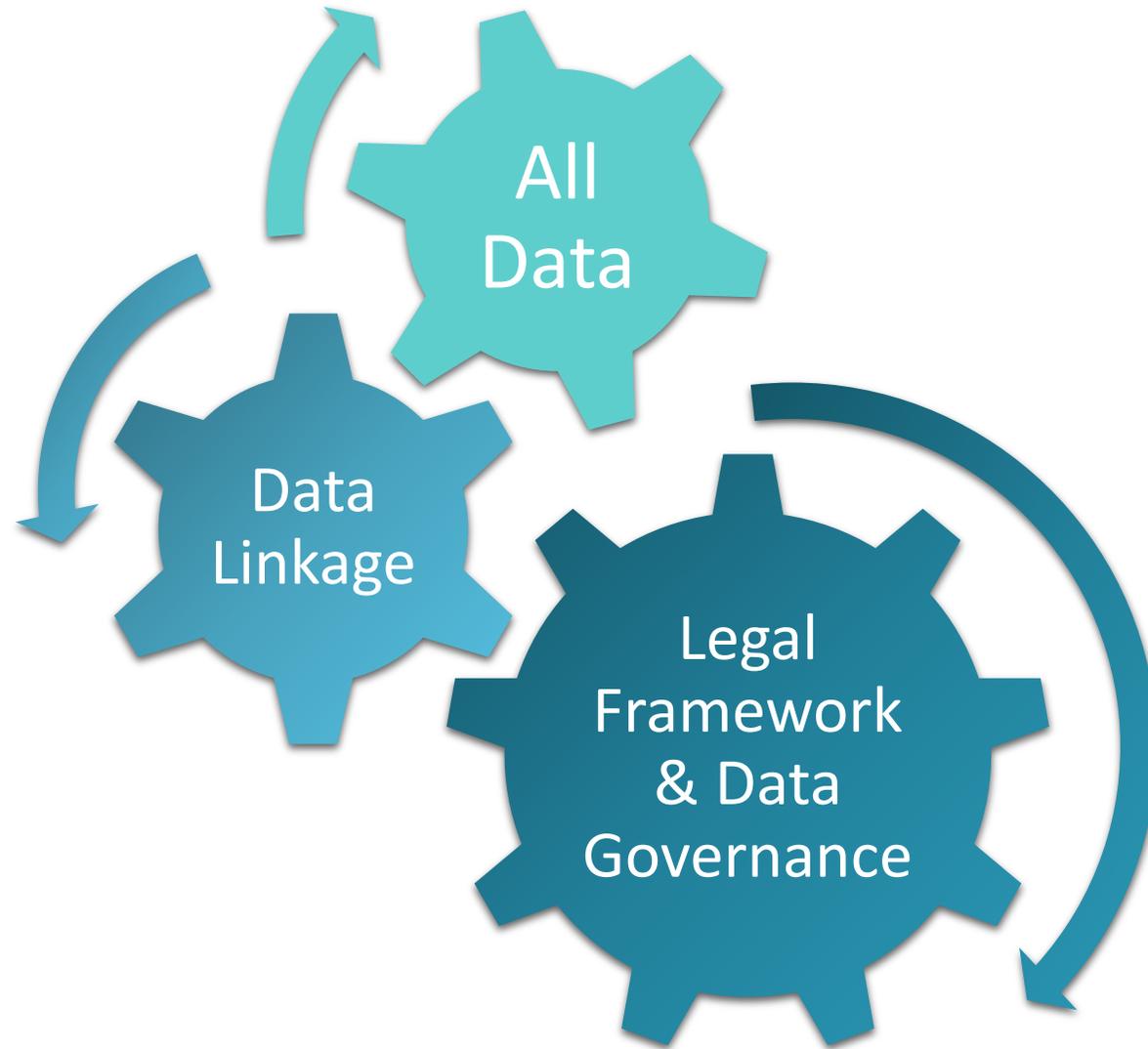
Exemplar: CRE for Heat



Illuminating the Ecosystem of Data Users to Elicit Purpose and Use



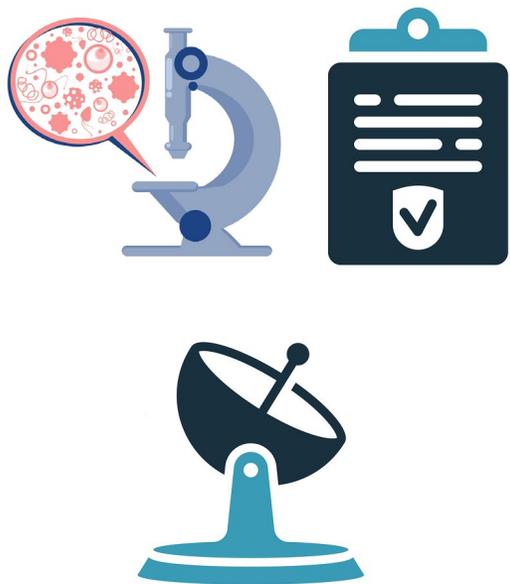
Leverage Data Infrastructure



Use *ALL* Data Assets

Going beyond the survey data we collect

Designed Data



Administrative Data



Opportunity Data



Procedural Data



Exploit Data Linkage

We collect a variety of linkage fields and tailor linkage methodology to each linkage type – people, places, jobs, and organizations.



Adhere to Legal Framework and Data Governance

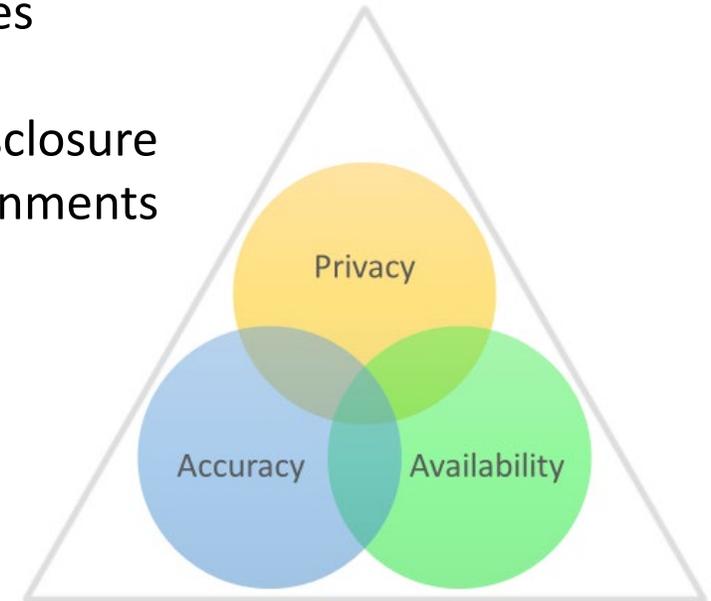
Title 13 directs us to acquire and use external data records for statistical purposes.

Foundations for Evidence-Based Policymaking Act promotes and encourages data sharing.

Confidential Information Protection and Statistical Efficiency Act sets forth functional separation of statistical versus administrative uses of data.

Ethical Data Stewardship

- ✓ Ensure only statistical uses
- ✓ Disclosure review
- ✓ Punishing of wrongful disclosure
- ✓ Secure computing environments

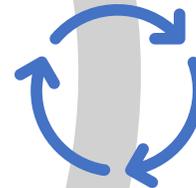


Statistical Product Spectrum for *ALL* Users

An opportunity
to identify and fill
the space in between



Protected
microdata
(in FSRDCs)



Simple, public
use products

Summing it up: A Curated Data Enterprise



Enabling Technologies for the Enterprise

Data Ingest and Collection for the Enterprise (*all* data assets)



Enterprise Data Lake

Data processing, computing, and management



Enterprise Linked Frames

4 seamless and linkable frames

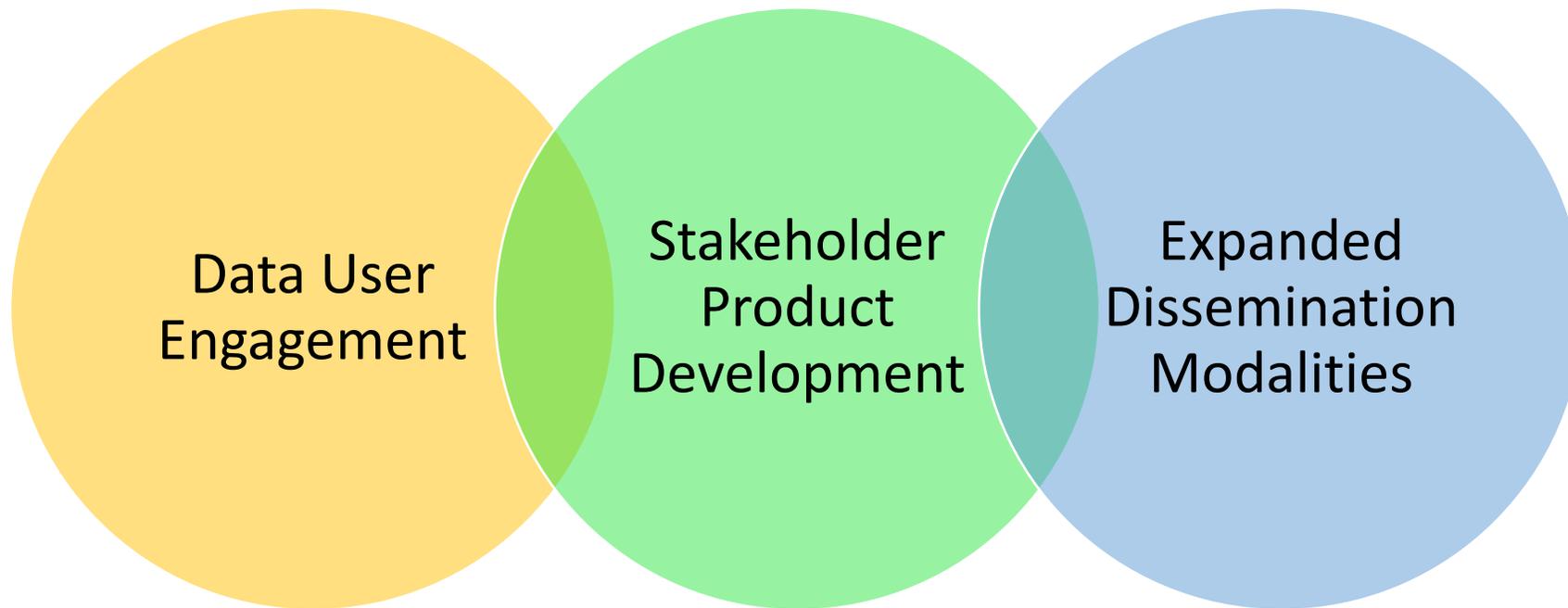


Enterprise Dissemination Services



Moving this Endeavor Forward

Our current research efforts for the Statistical Product First approach are focused around three core elements:



Statistical Product First Team Leads

**Erika Becker-Medina,
Data User Engagement**



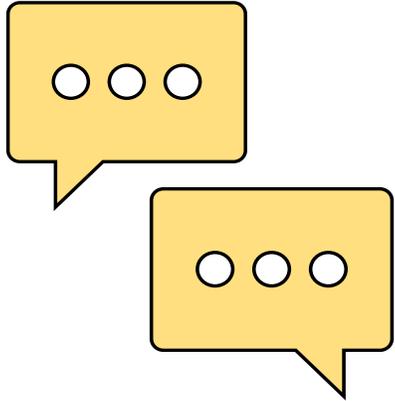
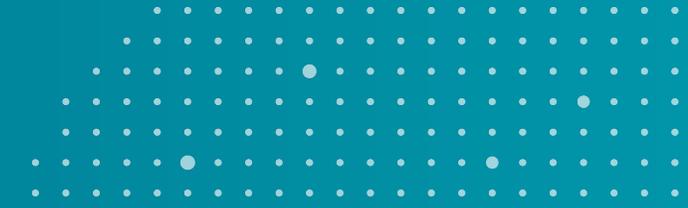
**Nate Ramsey,
Stakeholder Product Development**



**Emily Molfino,
Expanded Dissemination Modalities**



Data User Engagement Update



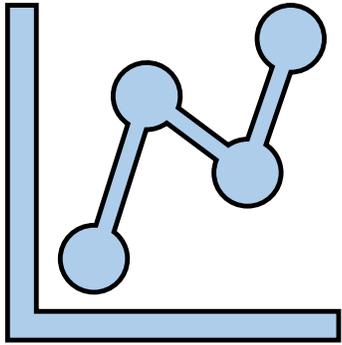
- Conducted multiple external engagement sessions and have heard similar feedback from different stakeholder groups: *Census does a lot of engagement already—it's the feedback loop that needs work!*
- Planning an internal data user engagement workshop for all Census data user outreach program representatives to map current processes, identify gaps, streamline disjointed efforts, and engage more effectively.
- Reviewing state data center sites to identify existing user-friendly data tools and take inventory of what state offices are doing to avoid redundancies and identify statistical product gaps.

Stakeholder Product Development Update

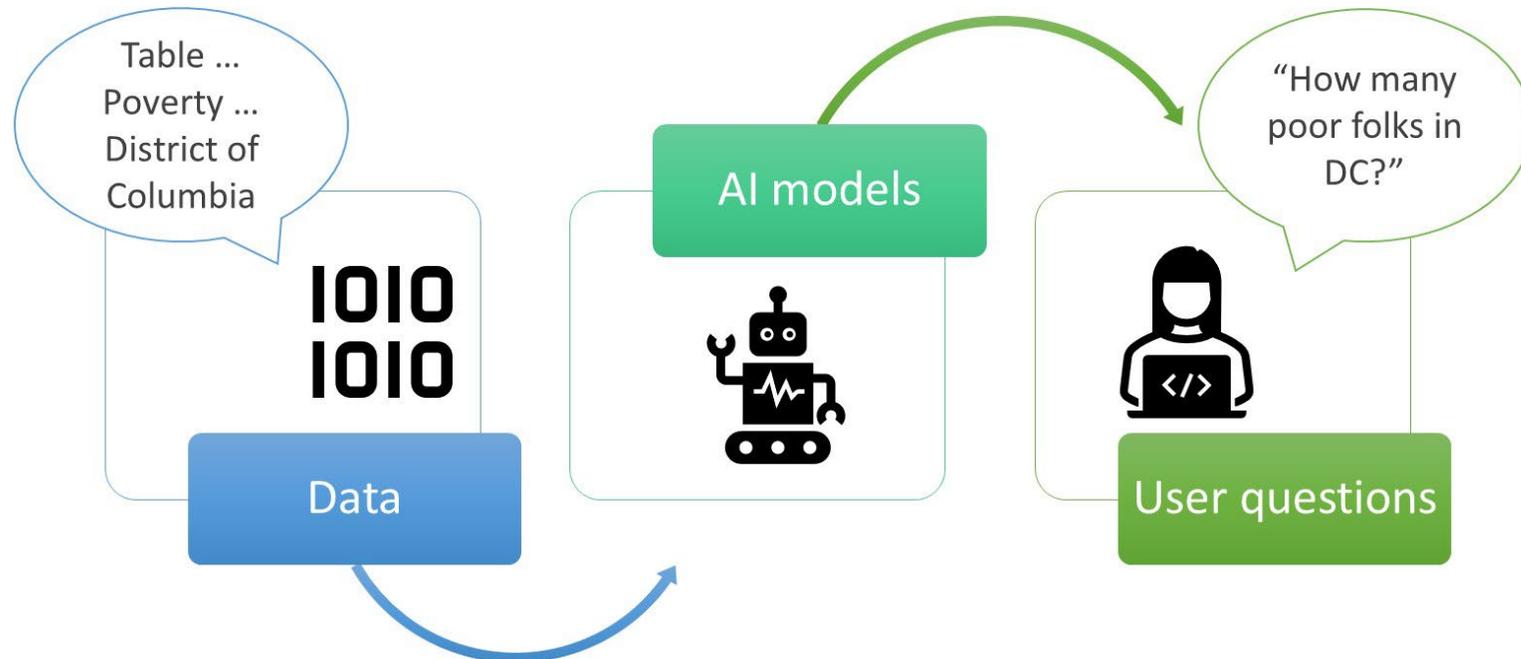


- Designing a demonstration pilot focused on a singular use case: identifying authoritative statistics to satisfy grant application requirements among state and local government and non-profit applicants.
- The demonstration pilot will encompass the three core elements: data user engagement, statistical product development, and expanded dissemination modalities.
- Similar strategies and techniques developed for this pilot could be applied toward a broader problem set (such as statistical needs among private sector organizations seeking government contracts.)

Expanded Dissemination Modalities Update



- Working with OpenAI and the Department of Commerce's Chief Data Officer to determine how to better deliver our data for use in AI platforms.
- Improving and expanding the contextual information about the data, putting it in a format that will be captured AI systems.





Imagine the Art of the Possible

Thank You

Sallie Ann Keller, Ph.D., Associate Director and Chief Scientist
Research and Methodology Directorate
U.S. Census Bureau

sallie.a.keller@census.gov