Total Survey Error through the Lens of Statistical Product First

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

Research and Methodology Directorate

U.S. Census Bureau

September 18, 2024

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

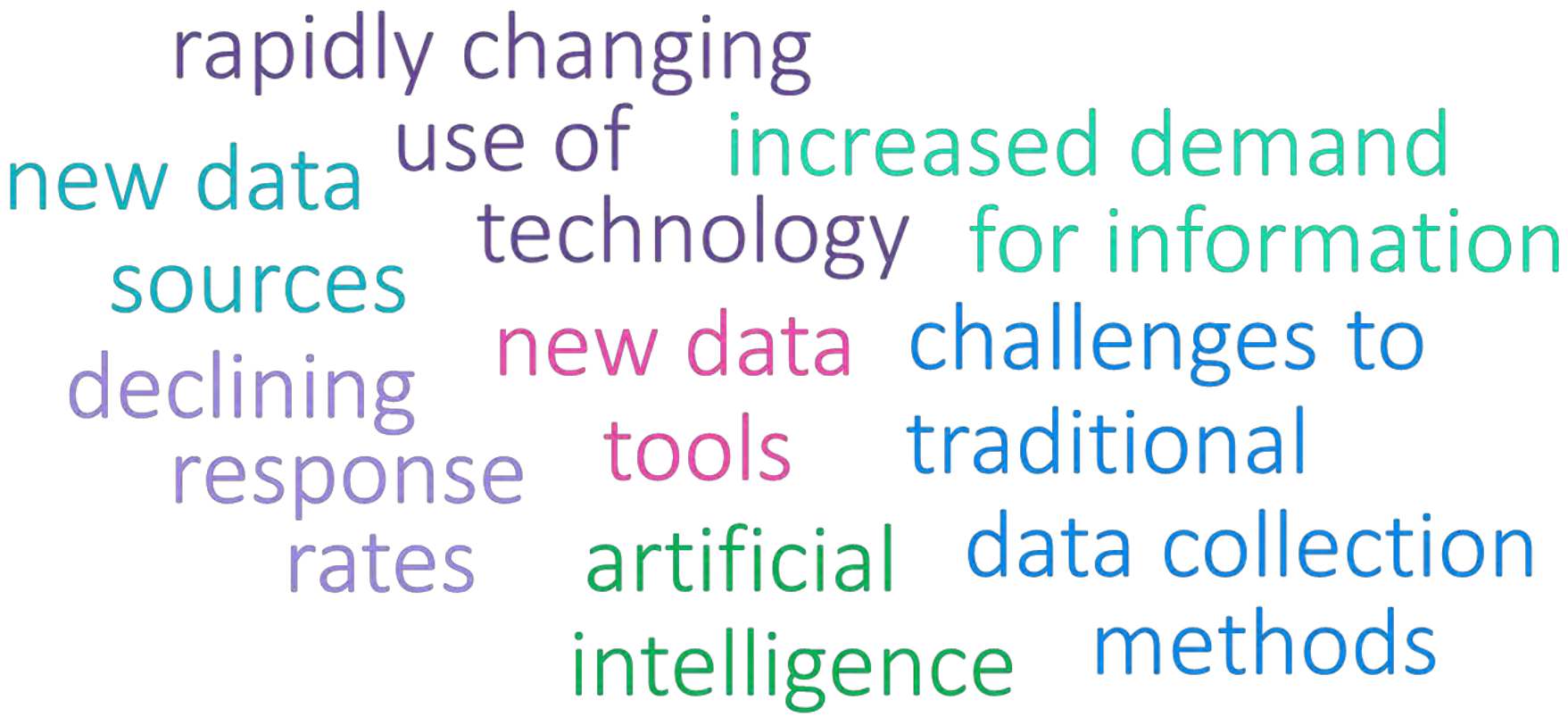
United States® Census Bureau

Acknowledgements:

Thanks to John Eltinge, Cass Dorius, Michael Hawes, and Erika Becker-Medina for their help in developing this presentation.

Also, thanks to Census Bureau staff and researchers for their contributions that led up to this effort.

Forces Driving Transformation Today



The Statistical Last Mile

United States Census Bureau

The Statistical Last Mile

20th Century

Federal government the dominant user

Statistical system a near monopoly

Output mostly cross-tabs

Published in books, mostly deposited in libraries, then electronically

Source data acquisition difficult and costly

Privacy and confidentiality risks were small

Computation expensive and limited

The Statistical Last Mile

20th Century	21st Century
Federal government the dominant user	Many diverse users
Statistical system a near monopoly	Many more organizations produce similar statistical products
Output mostly cross-tabs	Output varied and complex featuring visualizations and analysis tools
Published in books, mostly deposited in libraries, then electronically	Data accessed electronically online or in secure enclaves
Source data acquisition difficult and costly	Source data more abundant, available, and less costly—but not necessarily designed for statistical use
Privacy and confidentiality risks were small	Privacy and confidentiality risks are much greater
Computation expensive and limited	Computation vastly improved

The Statistical Last Mile Challenge

How does the Census Bureau deliver the right statistical information at the right time in actionable formats to address diverse data user needs?

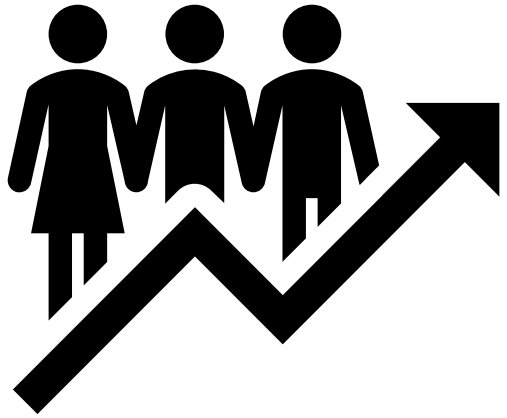
United States Census Bureau

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

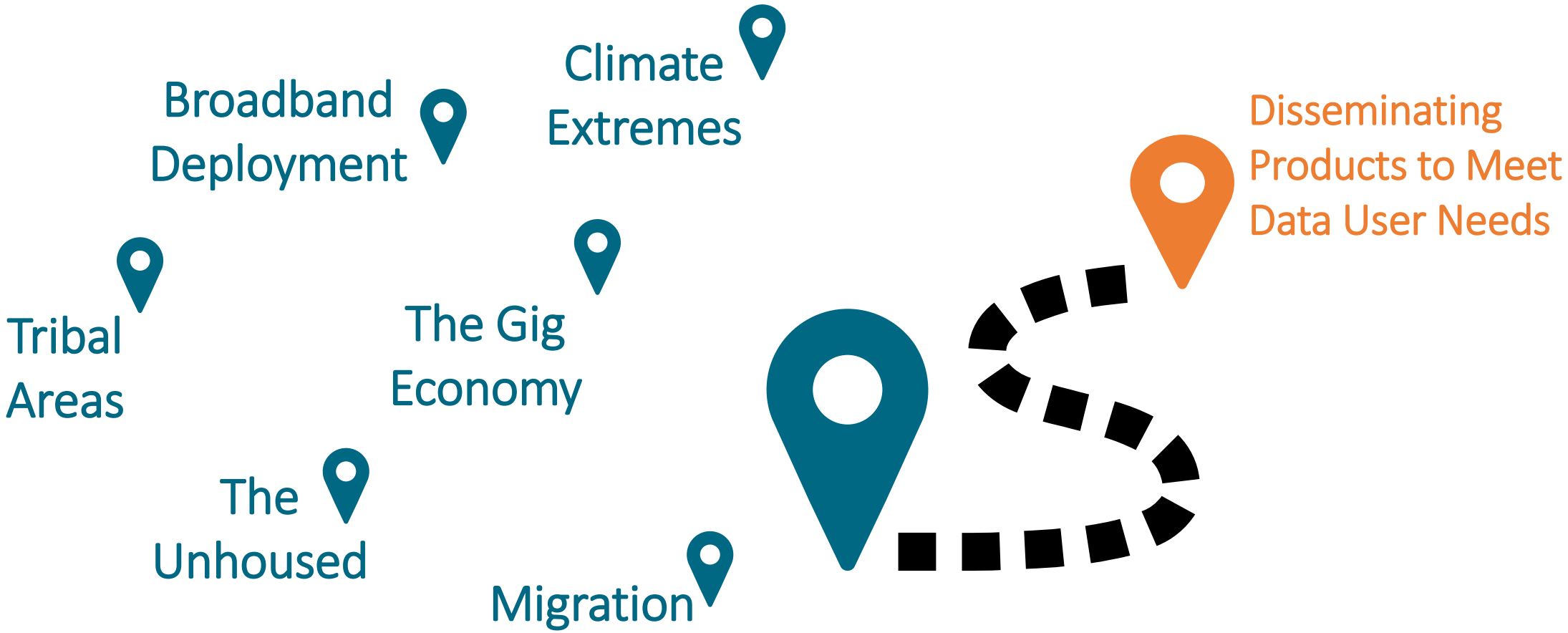


Moving from managing surveys to managing *statistics*

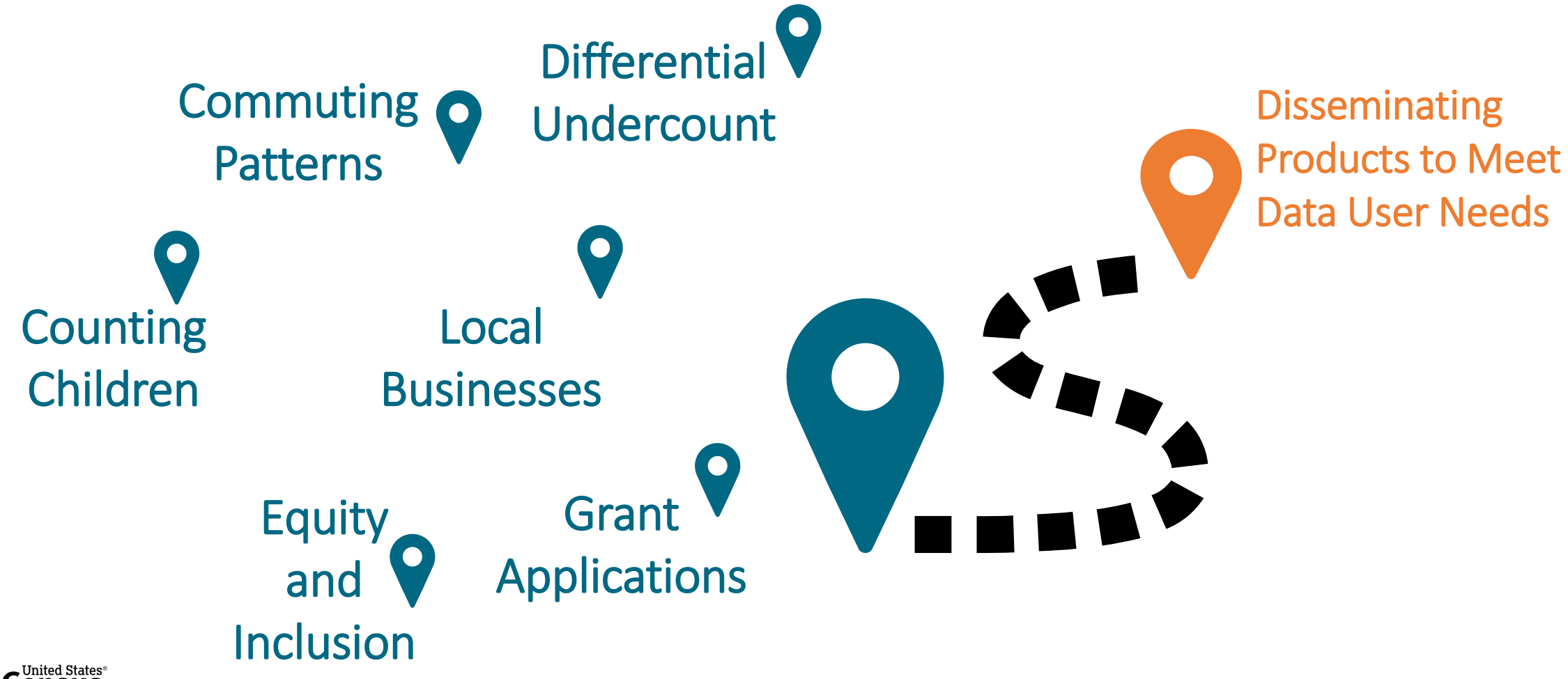
Statistical Product First Approach

Ensures Data Support Purposes and Uses

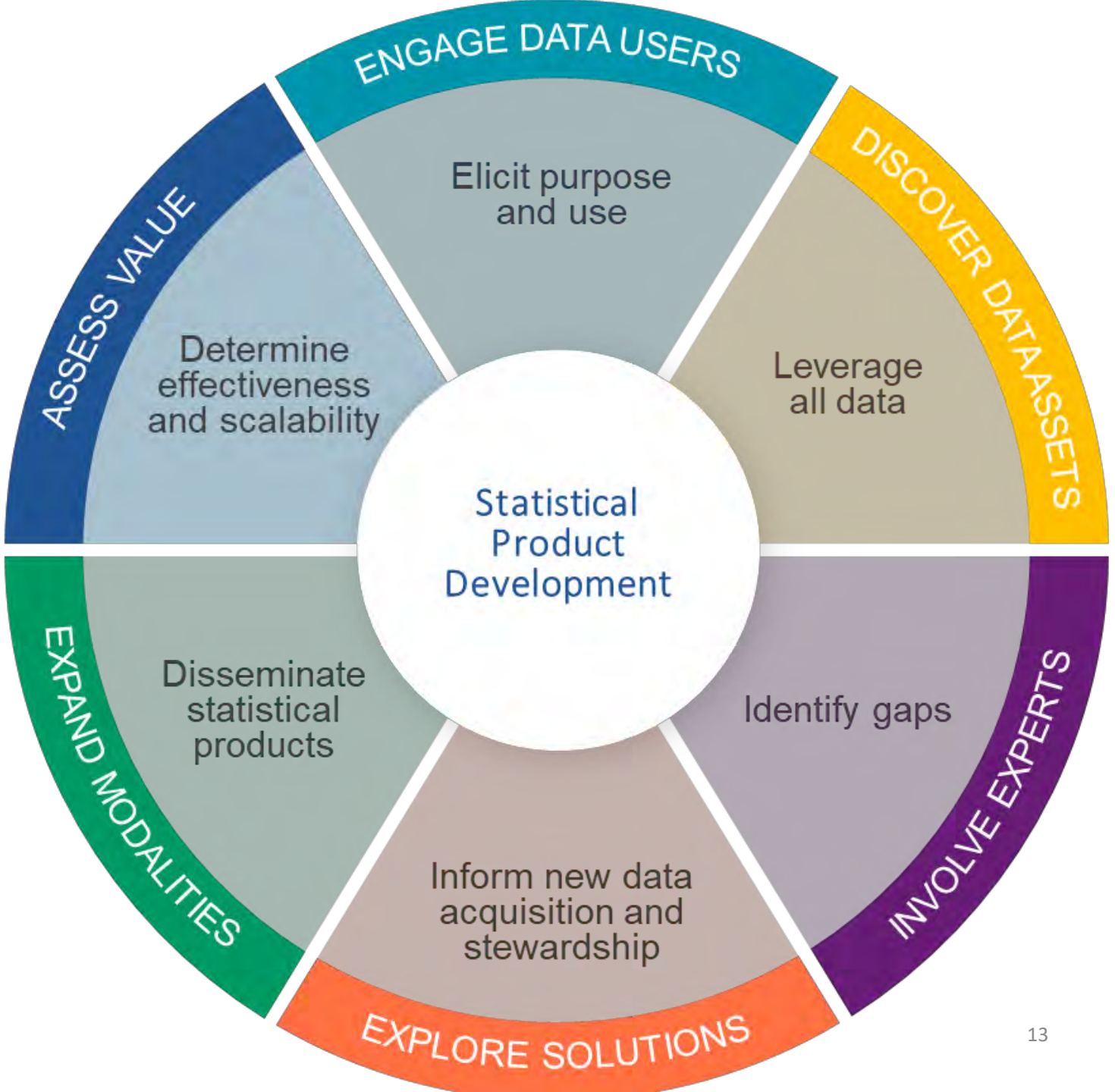
What are some purposes and uses?



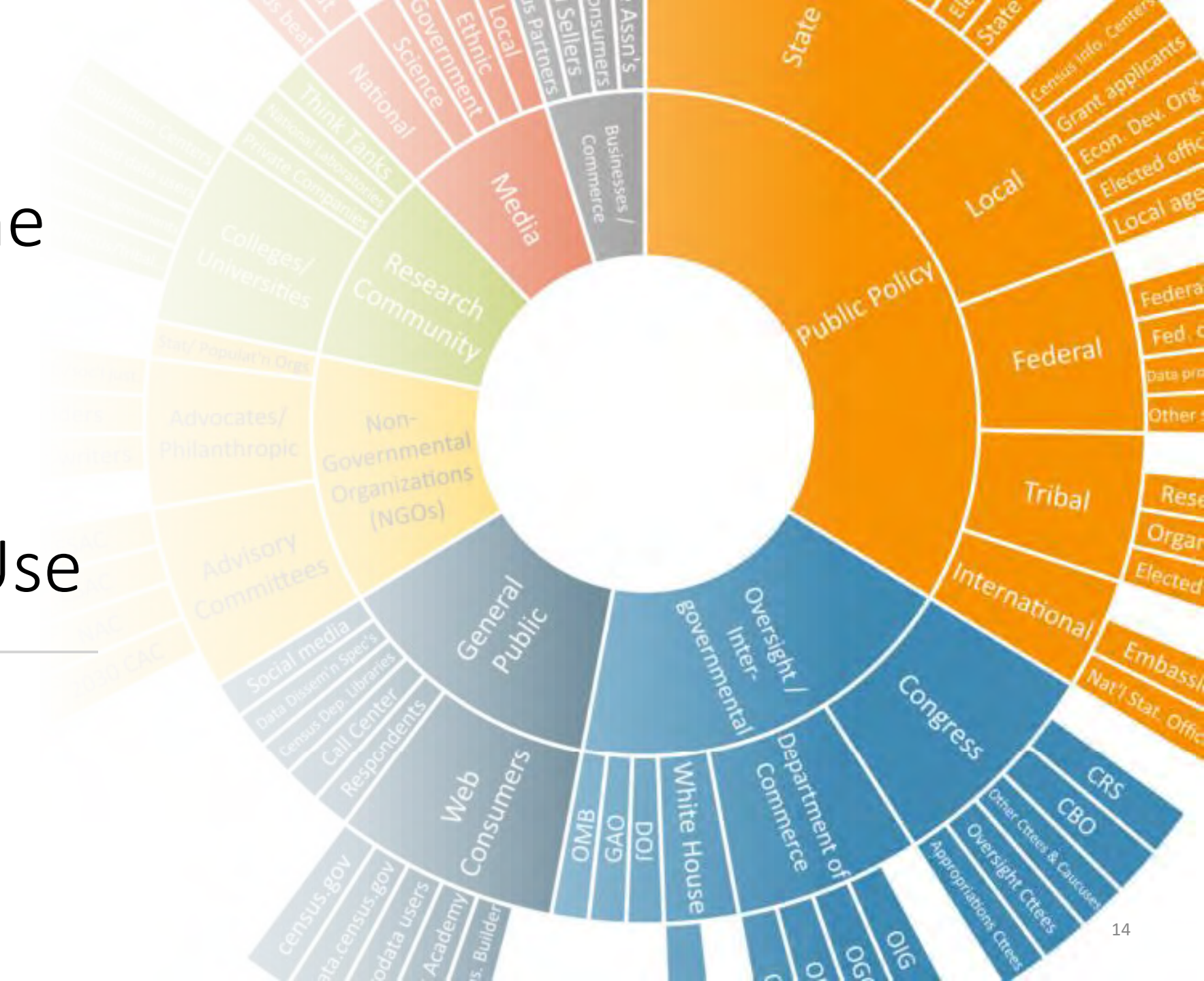
What are some purposes and uses?



Statistical Product First Innovation Cycle

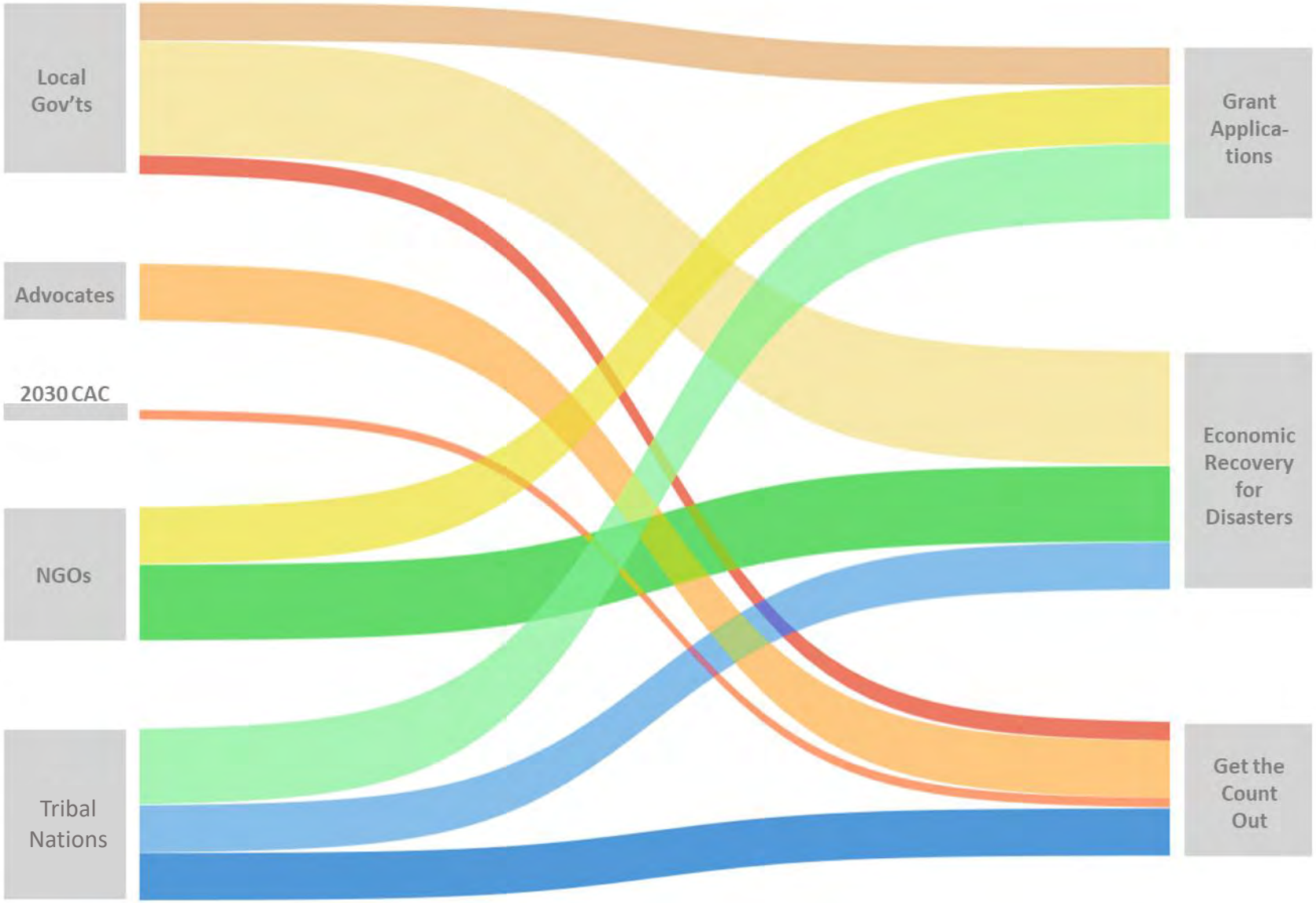


Illuminating the Ecosystem of Data Users to Elicit Purpose and Use



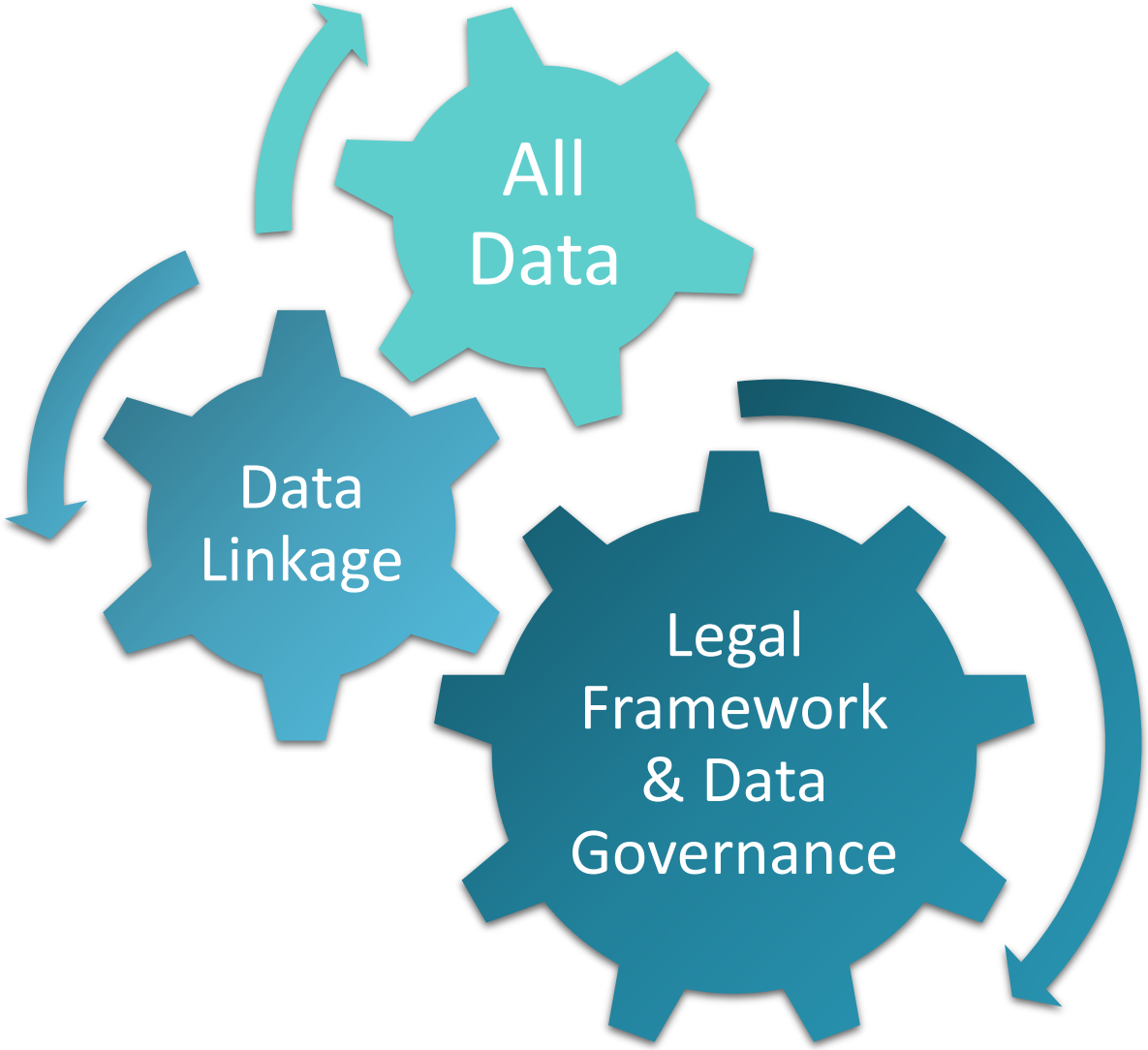
Develop User Segmentations

Currently developing a methodology to identify cross-cuts of the stakeholder group segmentation focused on common purpose and use needs.



United States Census Bureau

Leverage Data Infrastructure



United States Census Bureau

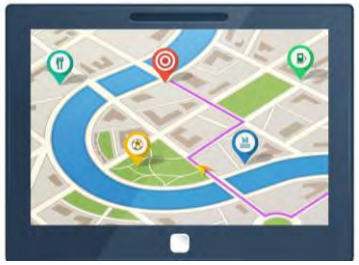
Use *ALL* Data Assets

Going beyond the survey data we collect

Designed Data

Administrative Data

Opportunity Data



Procedural Data

Harness Data Linkage

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

Statistical Product Spectrum for *ALL* Users

An opportunity
to identify and fill
the space in between

Protected microdata in Federal Statistical Research Data Centers (FSRDCs)

Simple, public use products

Adhere to Legal Framework and Data Governance

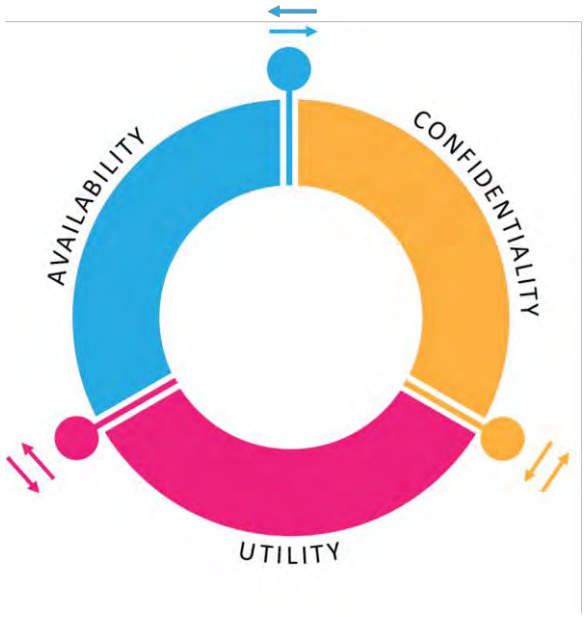
Title 13 *directs us to acquire and use external data records for statistical purposes.*

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

Responsible development and dissemination of Census Bureau products requires confidentially protection of data subjects' information while exclusively using this information for statistical purposes.



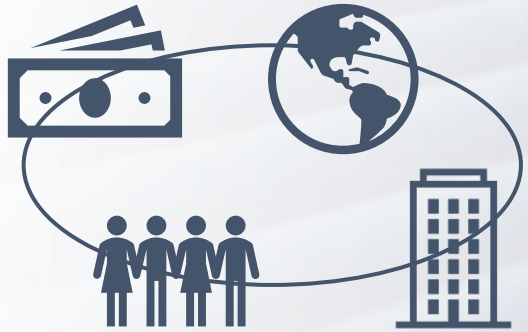
United States Census Bureau

Enabling Technologies

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

Enterprise Data Lake

Enterprise Linked Frames



Enterprise Dissemination Services

United States Census Bureau

Statistical Product First

Selected Demonstrations

Engagement with Indigenous Communities

National Indian Education Association: 54th Annual NIEA Convention & Trade Show

National Congress of American Indians: NCAI 80th Annual Convention & Marketplace

NCAI Executive Council Winter Session 2024

National Center for American Indian Enterprise Development (NCAIED) Reservation Economic Summit: RES 2024

Bureau of Indian Affairs: 2023 BIA Tribal Provider's Conference

U.S. Indigenous Data Sovereignty & Governance Summit 2024

NCAI 2024 Midyear Convention & Marketplace

United South and Eastern Tribes, Inc. and the USET Sovereignty Protection Fund: 2024 Impact Week

UPCOMING
Indigenous Data Sovereignty Symposium (sponsored by the Census Bureau)

United States® Census Bureau

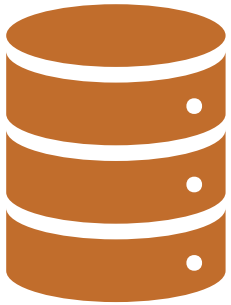
Collaboration through Engagement

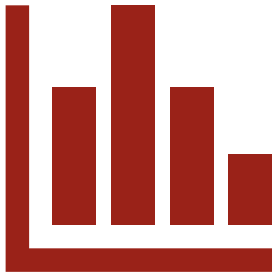
Improving Access to Tribal Data

Repackaging statistics across a variety of tribal and native boundaries through an infographic decision tree of current Census products. Collaborating on the dissemination with “My Tribal Area”.

Facilitating Administrative Records

Working with AD REC pilot project to help uncover tribal incentives to encourage and facilitate the provisioning of data to the Census Bureau.



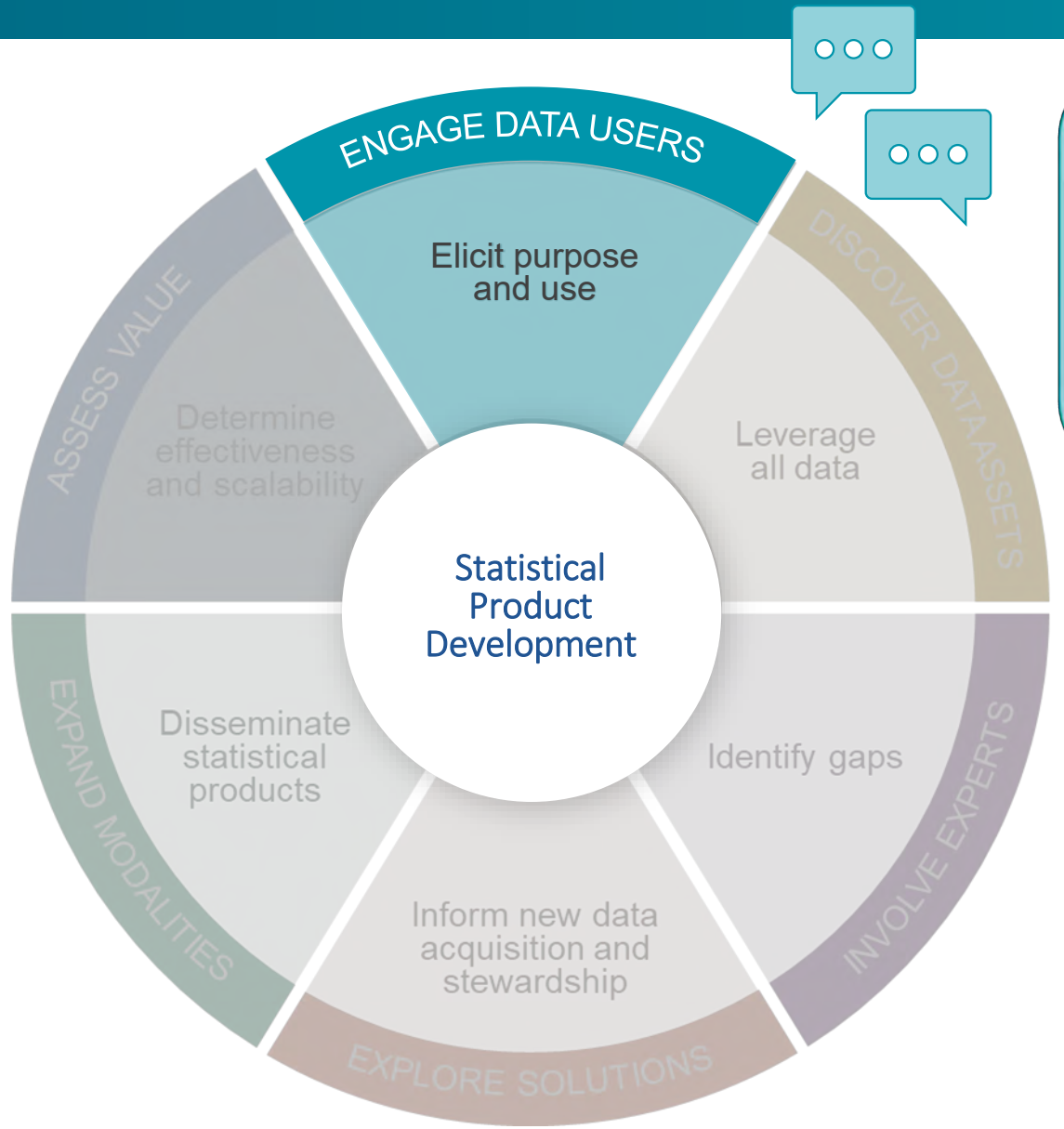


Exploring GDP for Tribal Areas

Exploring user feedback on the creation and value of GDP-like measures for tribal regions.

United States® Census Bureau

GDP-like Measure for Tribal Regions



Data user engagements led to powerful purpose and use needs:

We need a GDP-like product to measure the economic health of tribal regions.

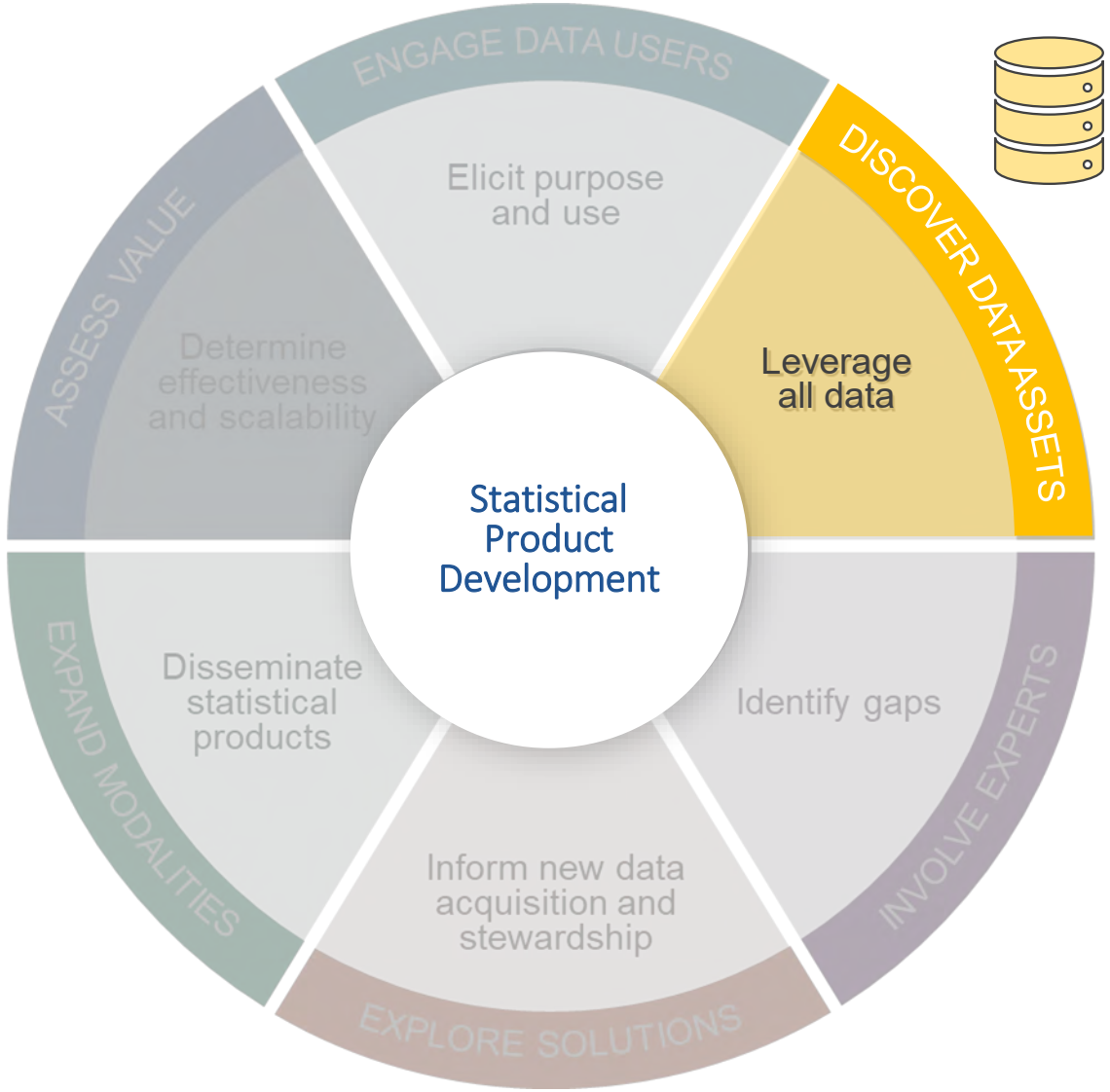
GDP-like Measure for Tribal Regions

Gross Domestic Product (GDP) is a common indicator used to measure a nation's economic health.

Data user engagements led to powerful purpose and use needs:

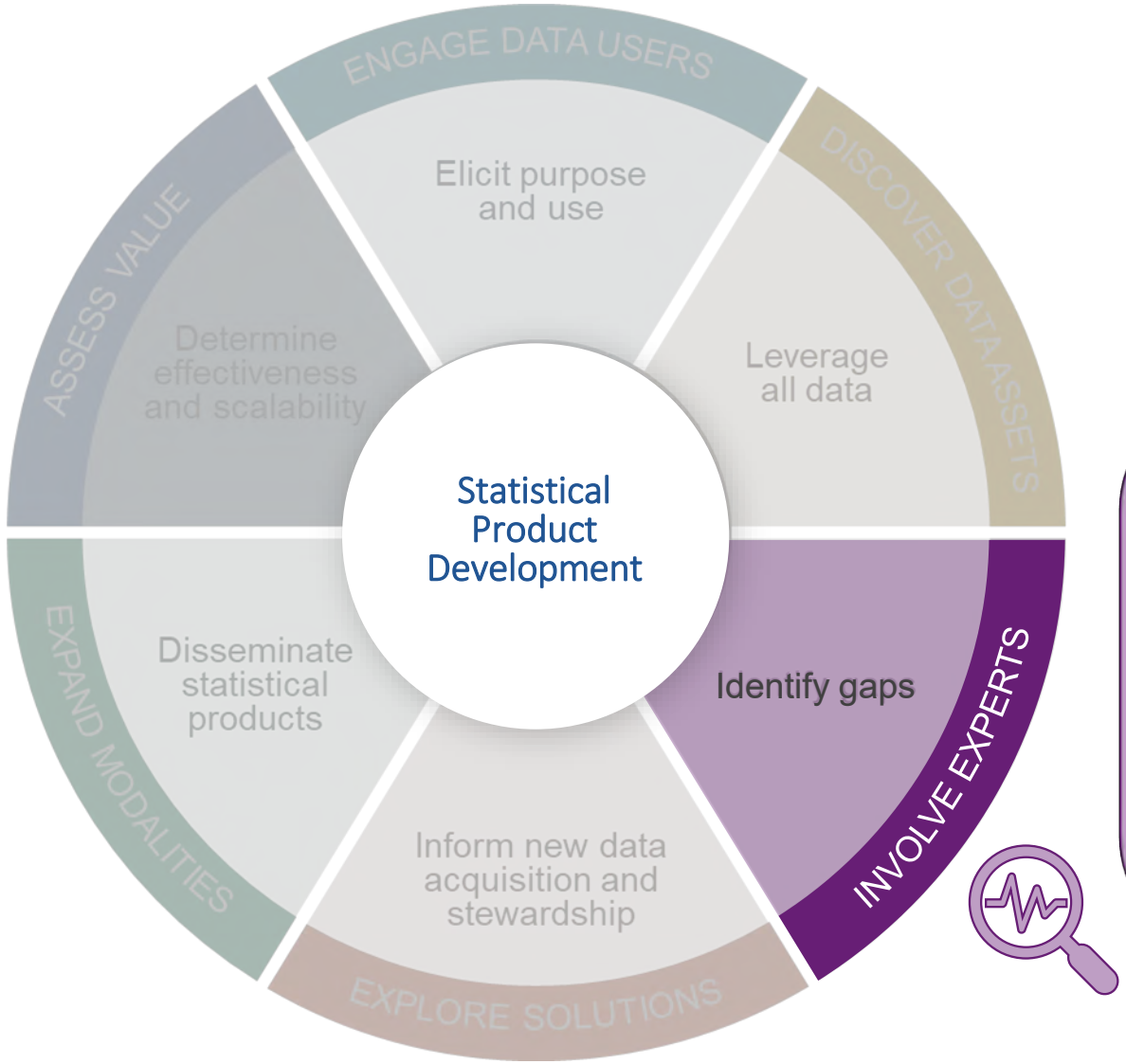
We need a GDP-like product to measure the economic health of tribal regions.

GDP-like Measure for Tribal Regions



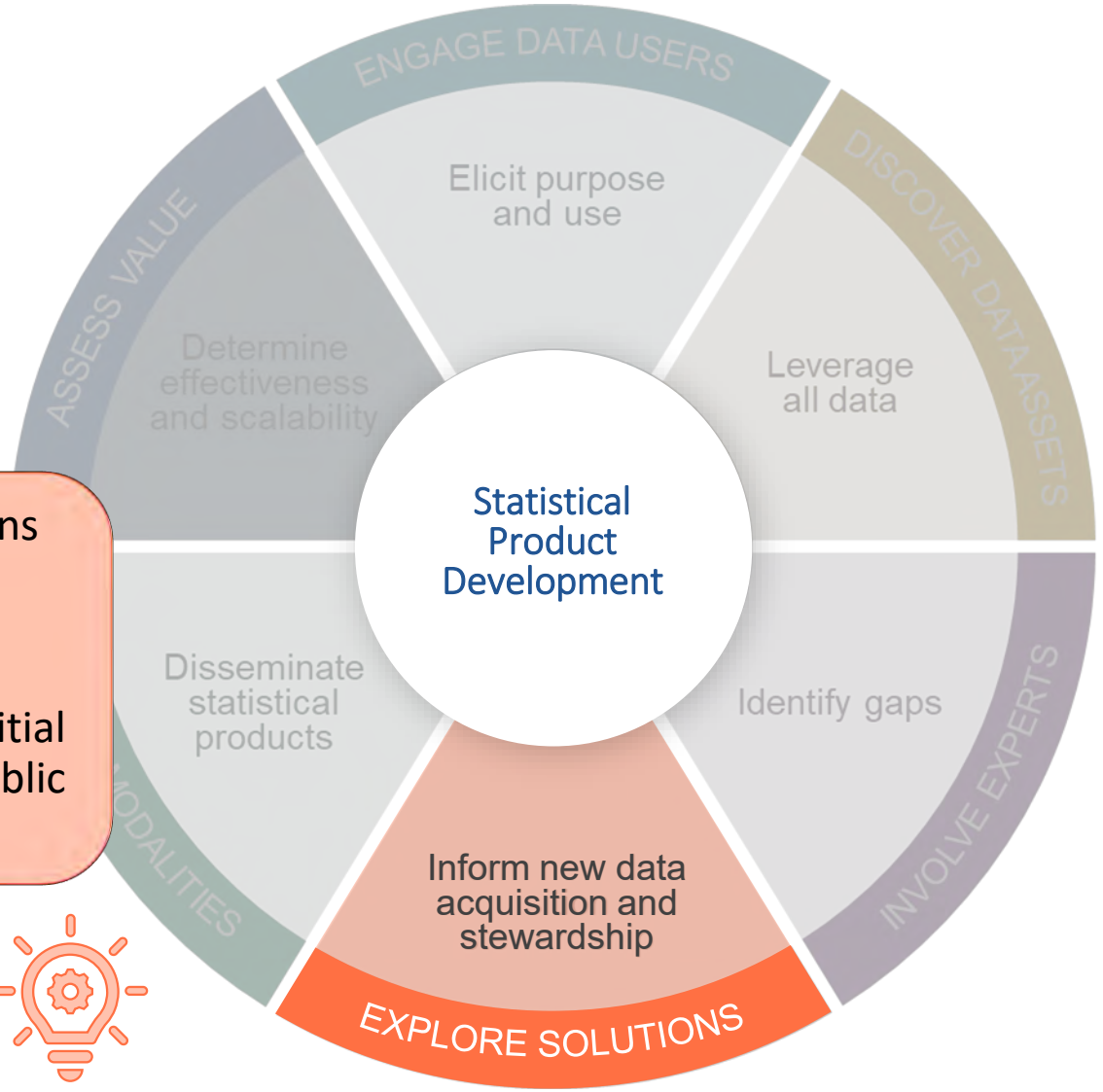
The Census Bureau collects vast amounts of economic data and then provides the estimates to the Bureau of Economic Analysis, where official GDP measures are produced.

GDP-like Measure for Tribal Regions



The GDP formula includes government spending. The Census Bureau collects this information as part of its Census of Governments. However, tribal nations are not included in that data collection.

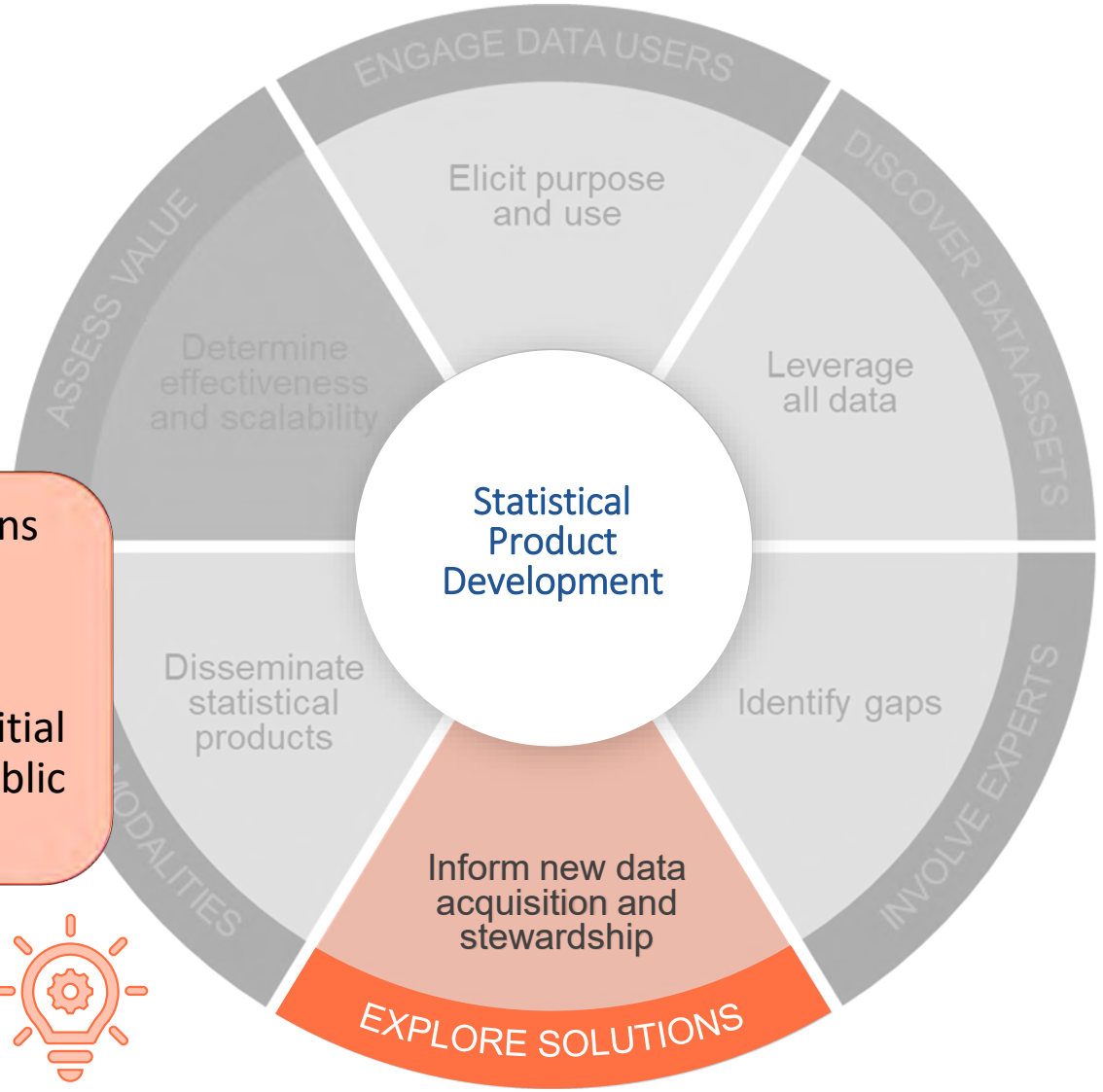
GDP-like Measure for Tribal Regions

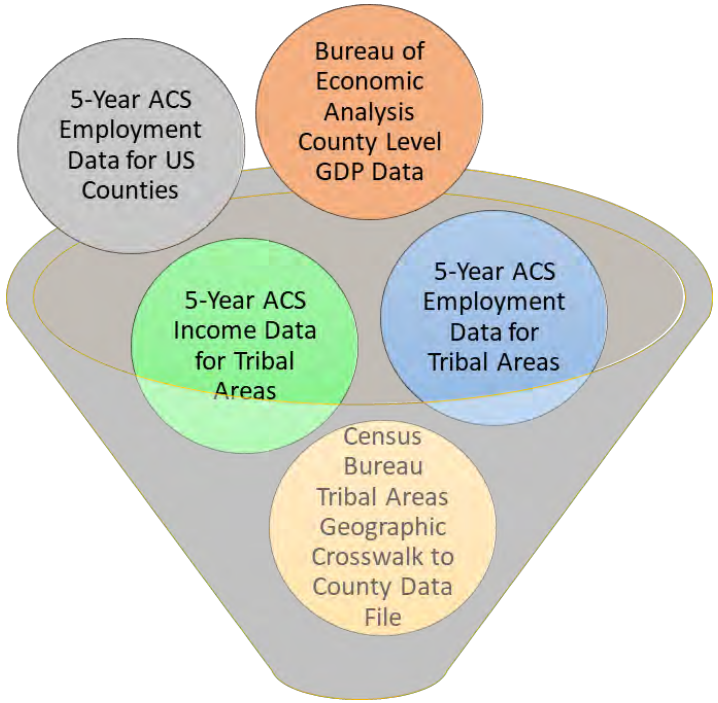


Currently exploring options for calculating a GDP-like measure but without the component of tribal government spending. Initial statistics are based on public use data sources.

GDP-like Measure for Tribal Regions

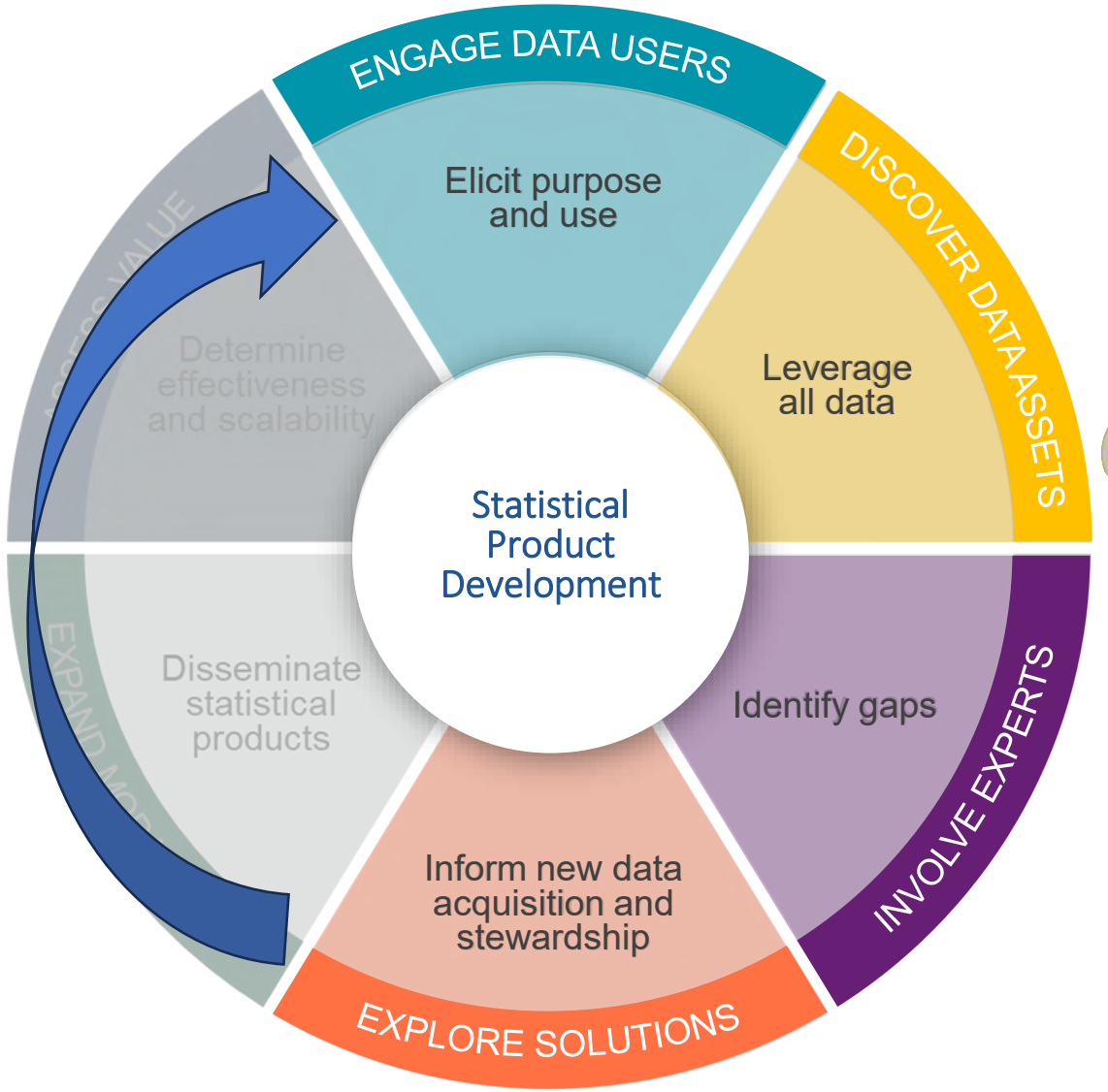
Currently exploring options for calculating a GDP-like measure but without the component of tribal government spending. Initial statistics are based on public use data sources.

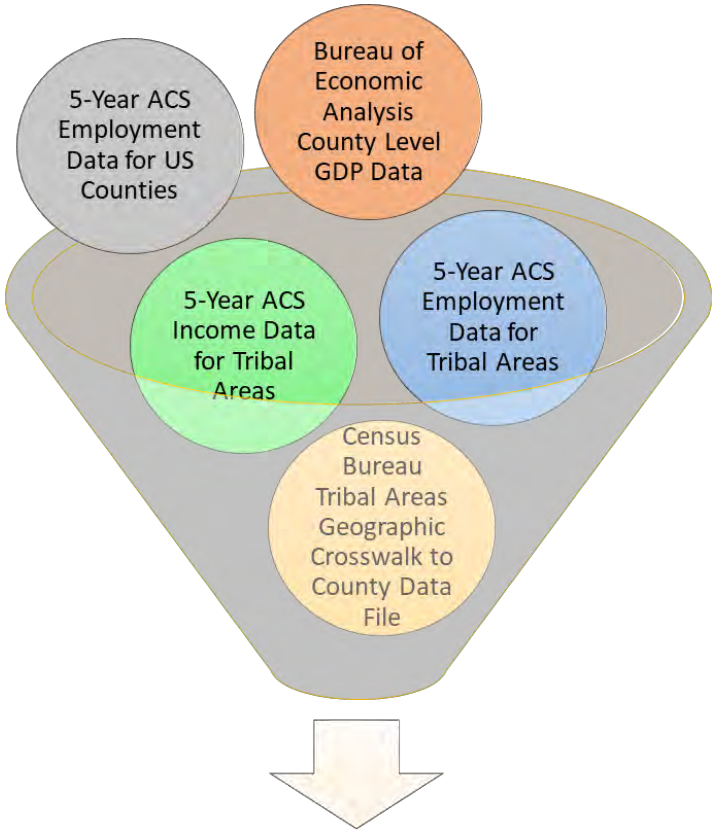




GDP-like Measure for Tribal Regions

Using early estimates to ground more engagements and discussions around purpose and use needs. These will inform more data discovery of confidential data assets and the refinement of new tribal GDP-like statistical products.





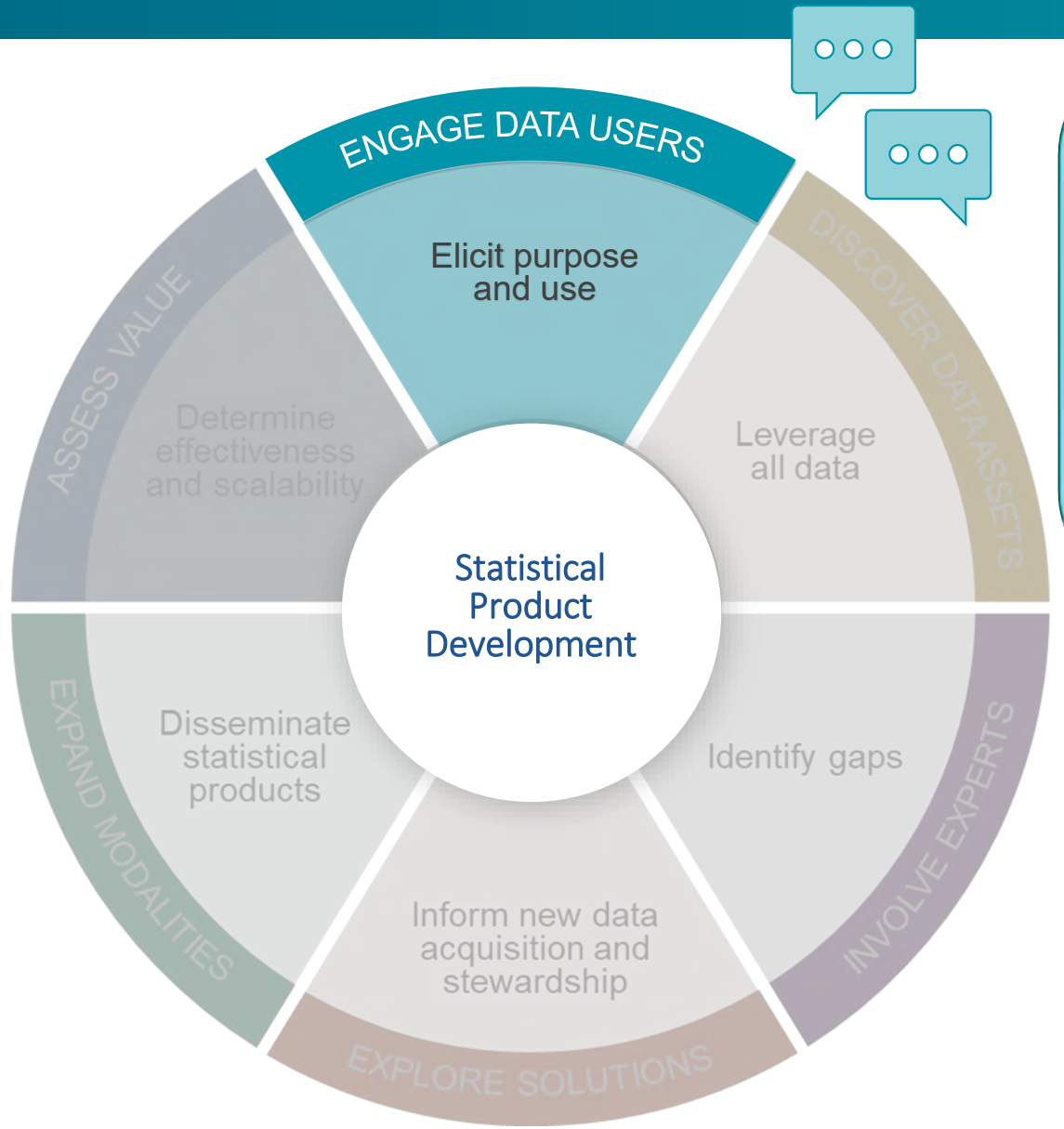
Initial GDP-like estimates for review and discussion

Seeking to support the needs of grant applicants

State and local governments, non-profits, and tribal organization members (SLaNT) have all shared the need for easier access to Census data to support funding applications on Grants.gov.

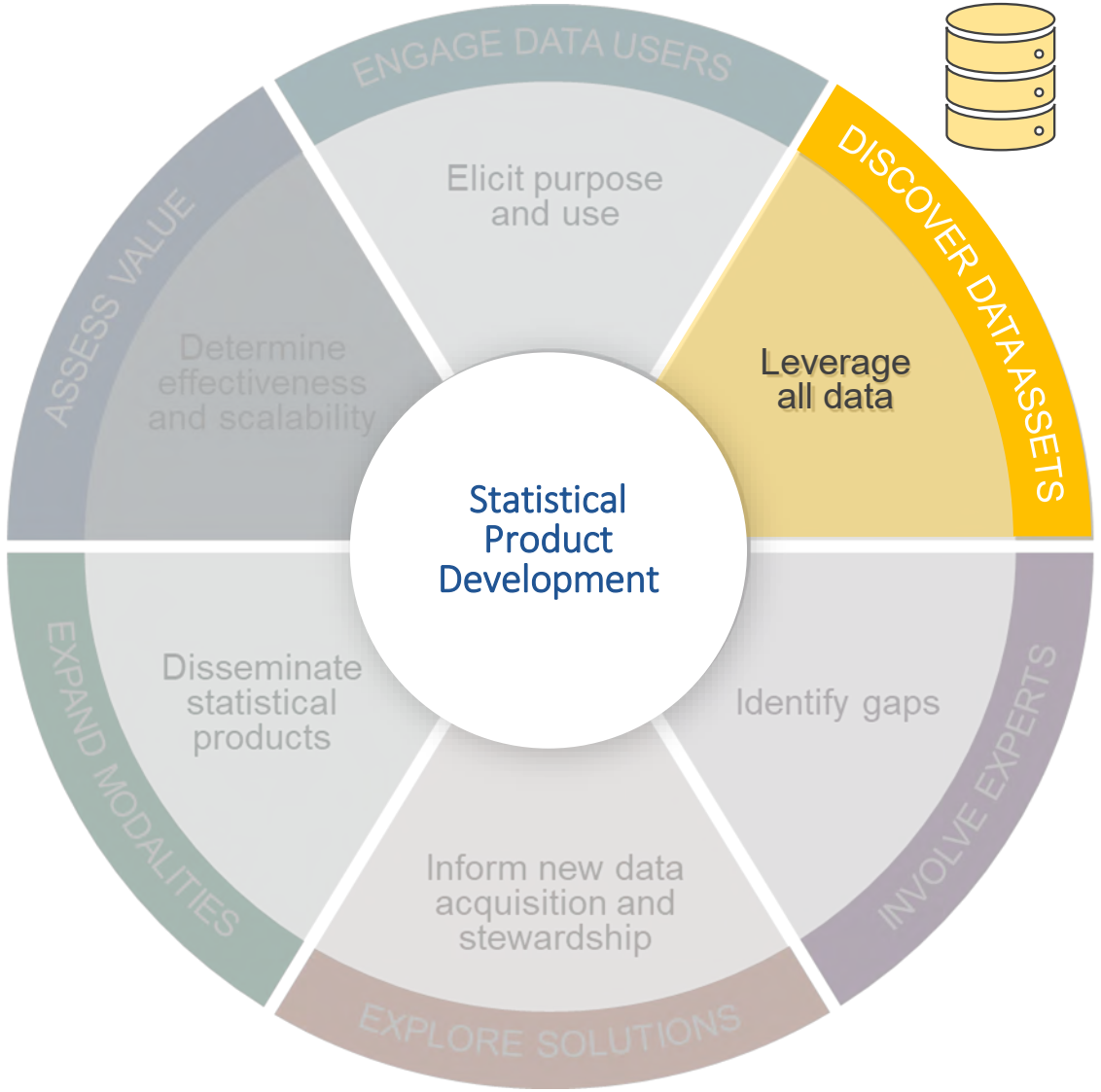
GRANTS.GOV℠
FIND. APPLY. SUCCEED.®

Supporting Grants.gov Applicants



We compiled information from listening sessions to create four personas that reflected stakeholders' needs and could be used to help support the goals of state and local governments, non-profits, and tribal organization members (SLaNT).

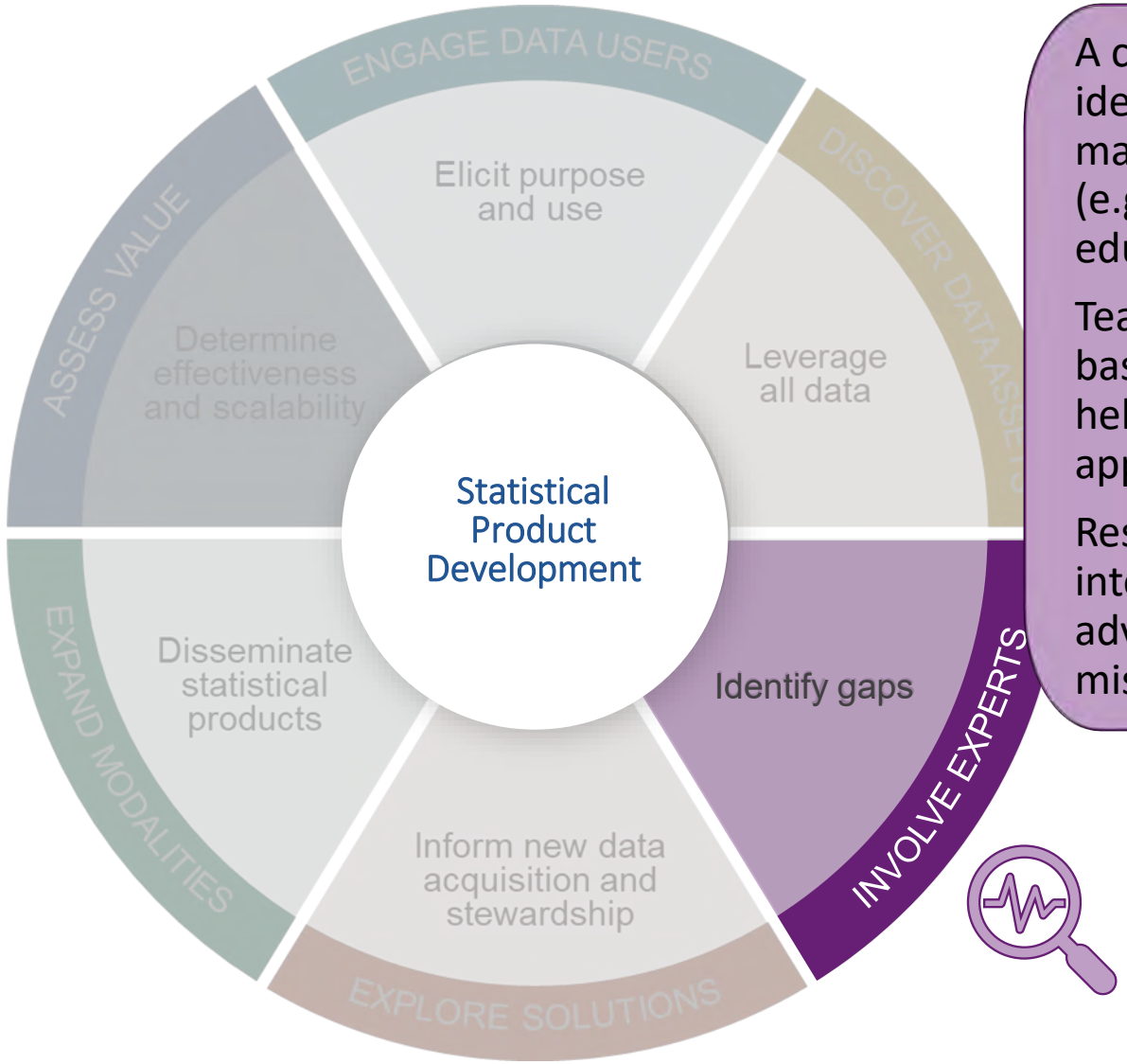
Supporting Grants.gov Applicants



We analyzed all grants currently available on Grants.gov and identified the frequency and availability of funding opportunities by data domain.

Then, we identified potential Census data sources that support applications for the known funding opportunities.

Supporting Grants.gov Applicants



A criteria matrix was used to identify which data domains matter most to our stakeholders (e.g., health and nutrition, education, transportation, etc.).

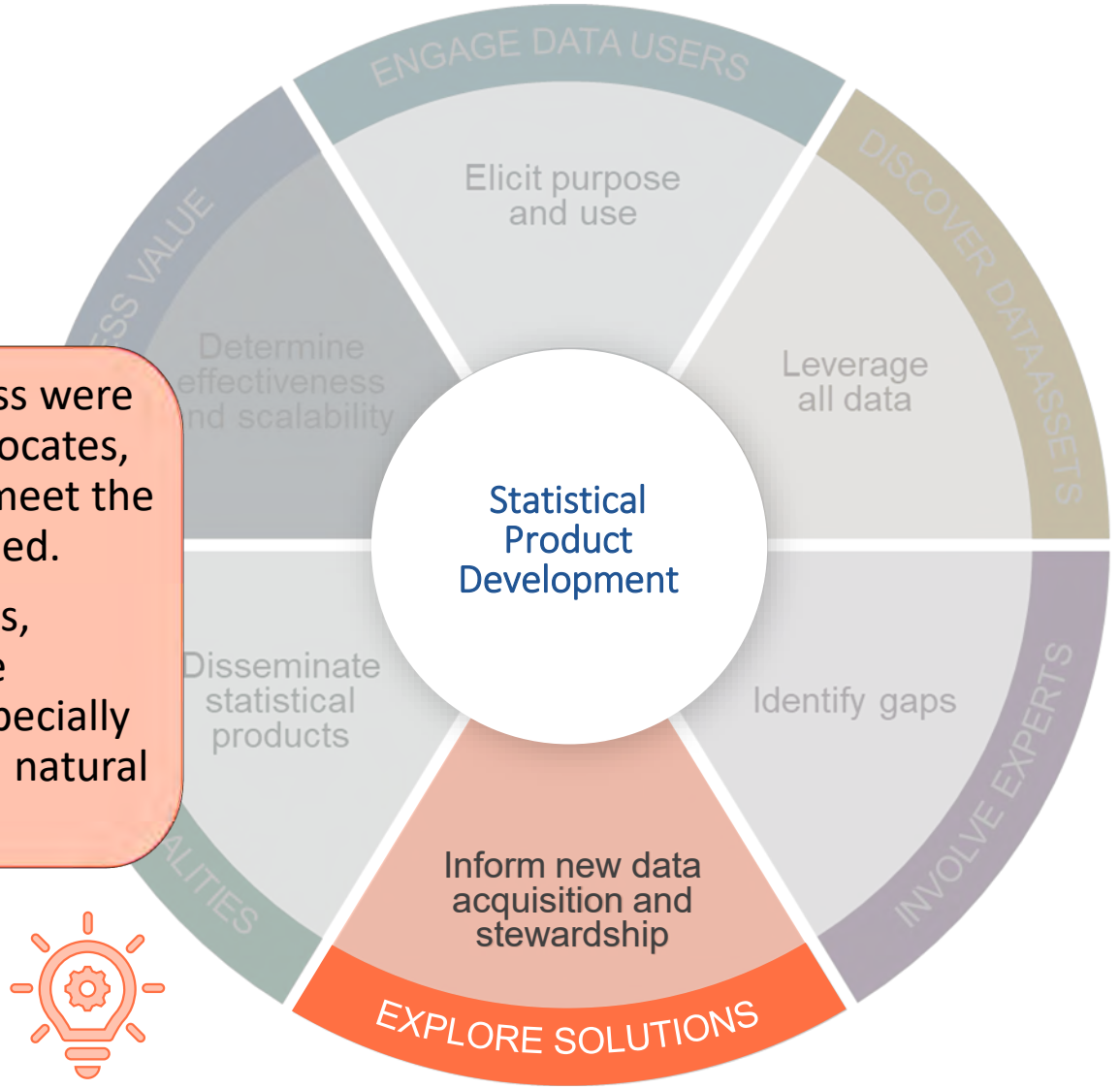
Teams identified a list of 100 basic data elements that would help diverse SLaNT data users apply for grants.

Results were shared back with internal experts and user advocates to identify what was missing.

Supporting Grants.gov Applicants

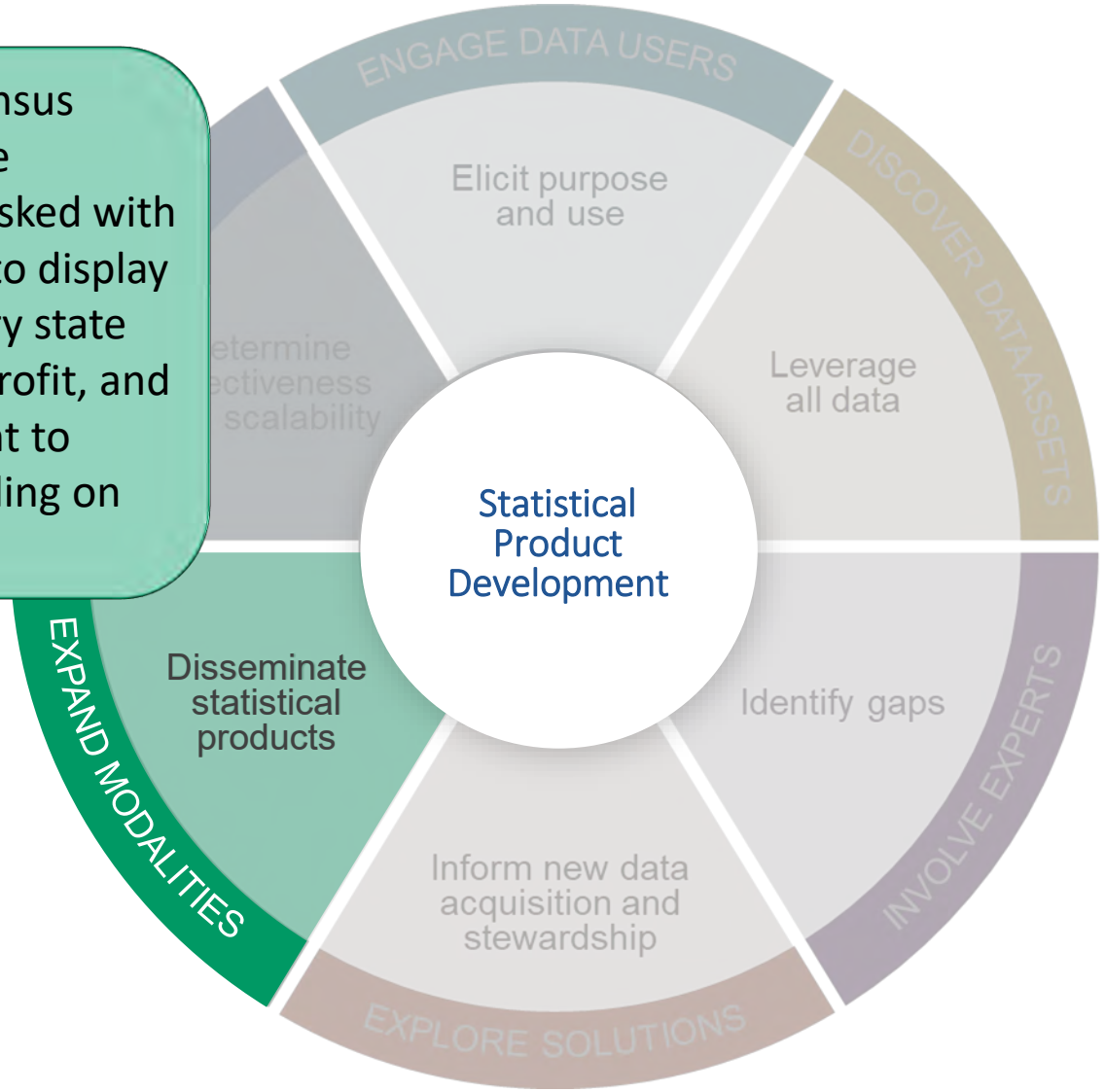
Granularity and timeliness were key concerns of user advocates, so data was selected to meet the screening criteria identified.

We identified non-Census, public data that might be needed to fill in gaps, especially around environment and natural resources.



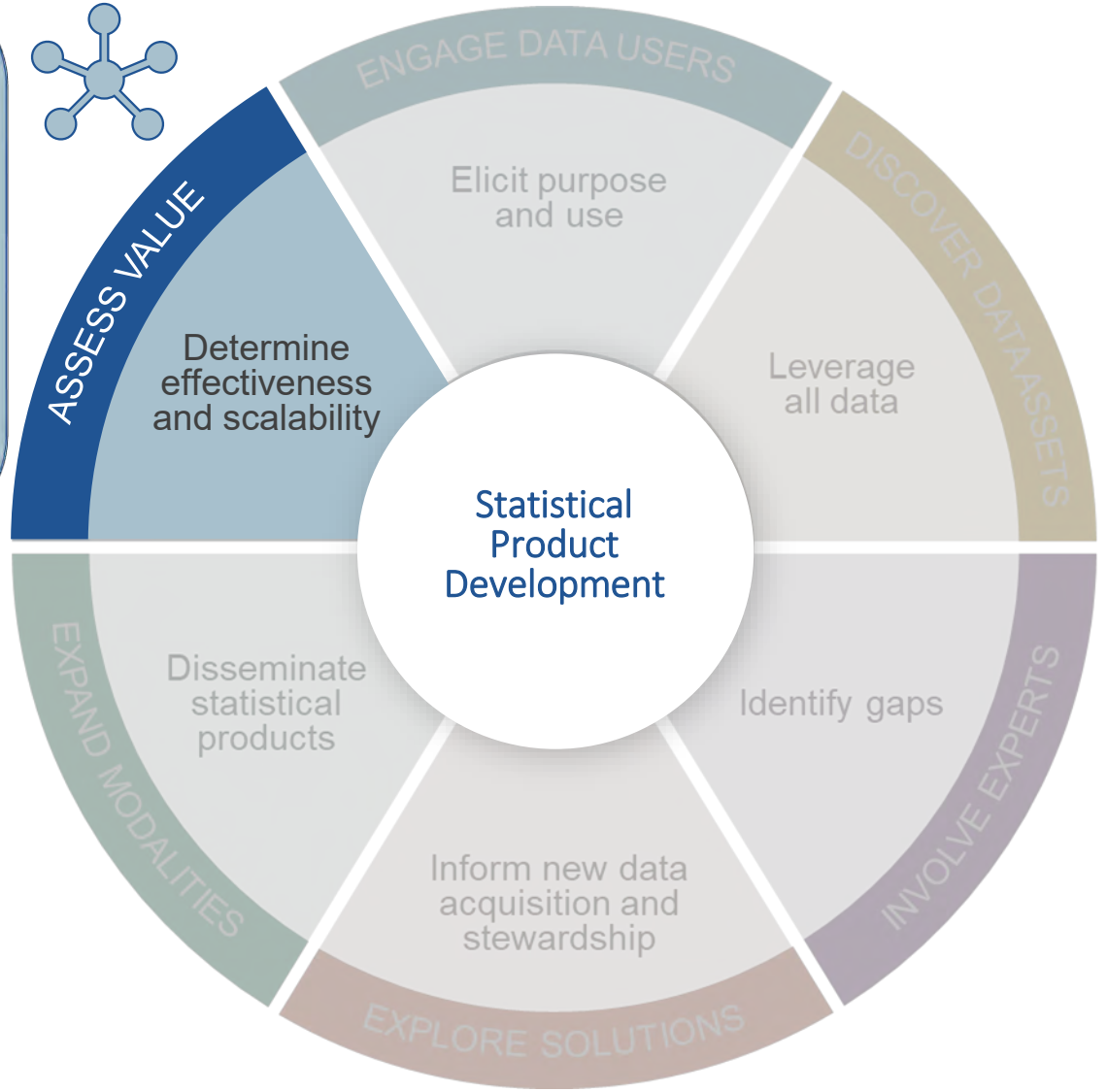
Supporting Grants.gov Applicants

Capstone teams from the Census Bureau, Georgetown, and the University of Virginia were tasked with developing a prototype tool to display the 'Top 100' data items every state and local government, non-profit, and tribal organization might want to know when applying for funding on Grants.gov.



Supporting Grants.gov Applicants

Plans are underway to share the ‘Top 100’ items and tool prototype with stakeholders and user advocates to identify what works and what doesn’t before this tool goes through a second innovation cycle on its way to becoming a public facing statistical product.



How do we ensure quality—specifically, *data **and** statistical product quality*—is captured throughout the SPF Innovation Cycle?

United States® Census Bureau

Reflecting on the Evolution of Data Quality

Data quality is a relative concept, consequently metrics change depending on use.

ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION Volume 4, 2017

Review Article | Free

The Evolution of Data Quality: Understanding the Transdisciplinary Origins of Data Quality Concepts and Approaches

Sallie Keller[1], Gizem Korkmaz[1], Mark Orr[1], Aaron Schroeder[1], and Stephanie Shipp[1]

View Affiliations

Vol. 4:85-108 (Volume publication date March 2017) | https://doi.org/10.1146/annurev-statistics-060116-054114

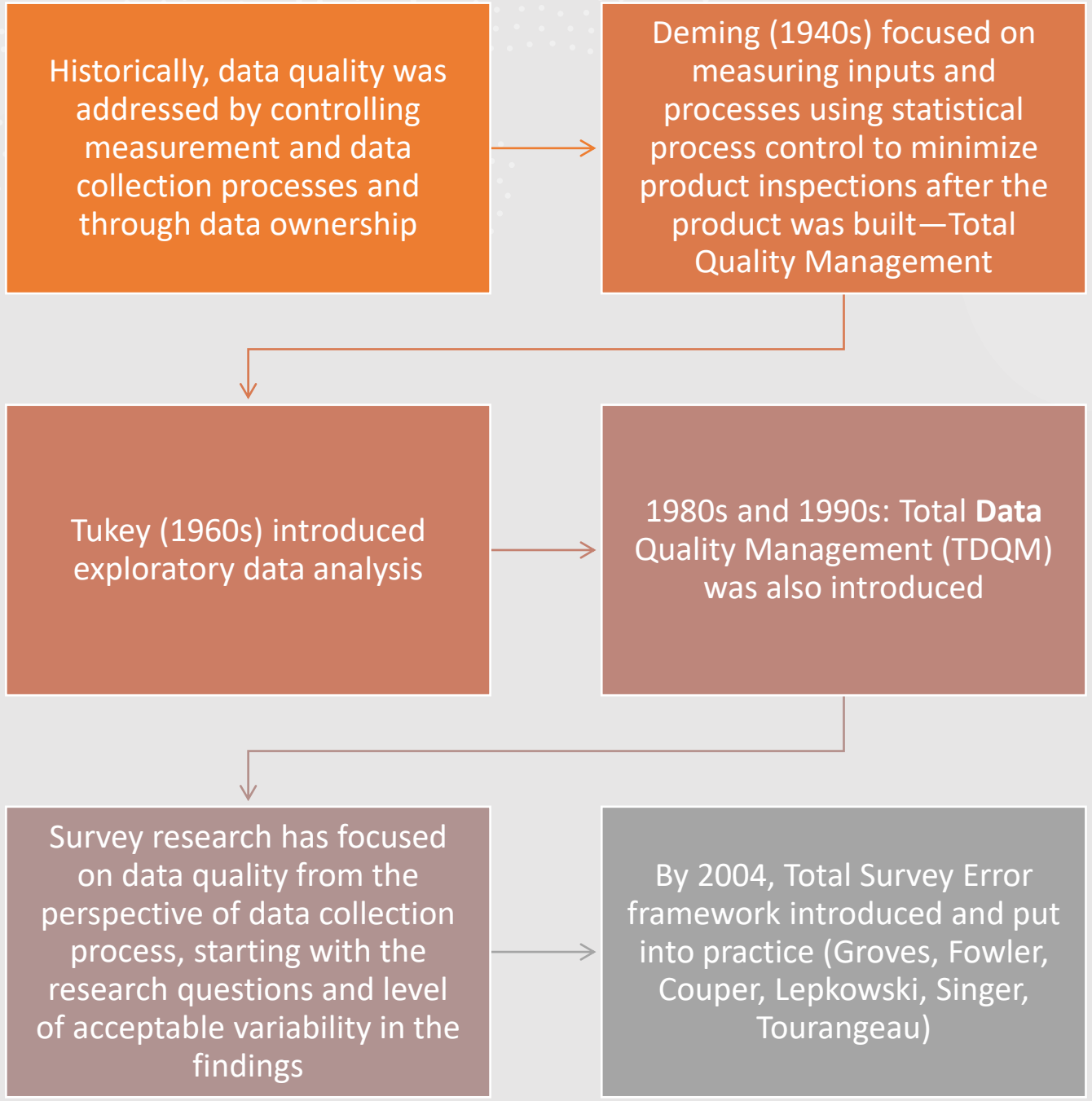
First published as a Review in Advance on January 06, 2017

“Far better an approximate answer to the right questions, which is often vague, than an exact answer to the wrong questions, which can always be made precise.”

~ John Tukey

Reflecting on the Evolution of Data Quality

Statistics—as the science of uncertainty—has been central to data quality across all fields of endeavor



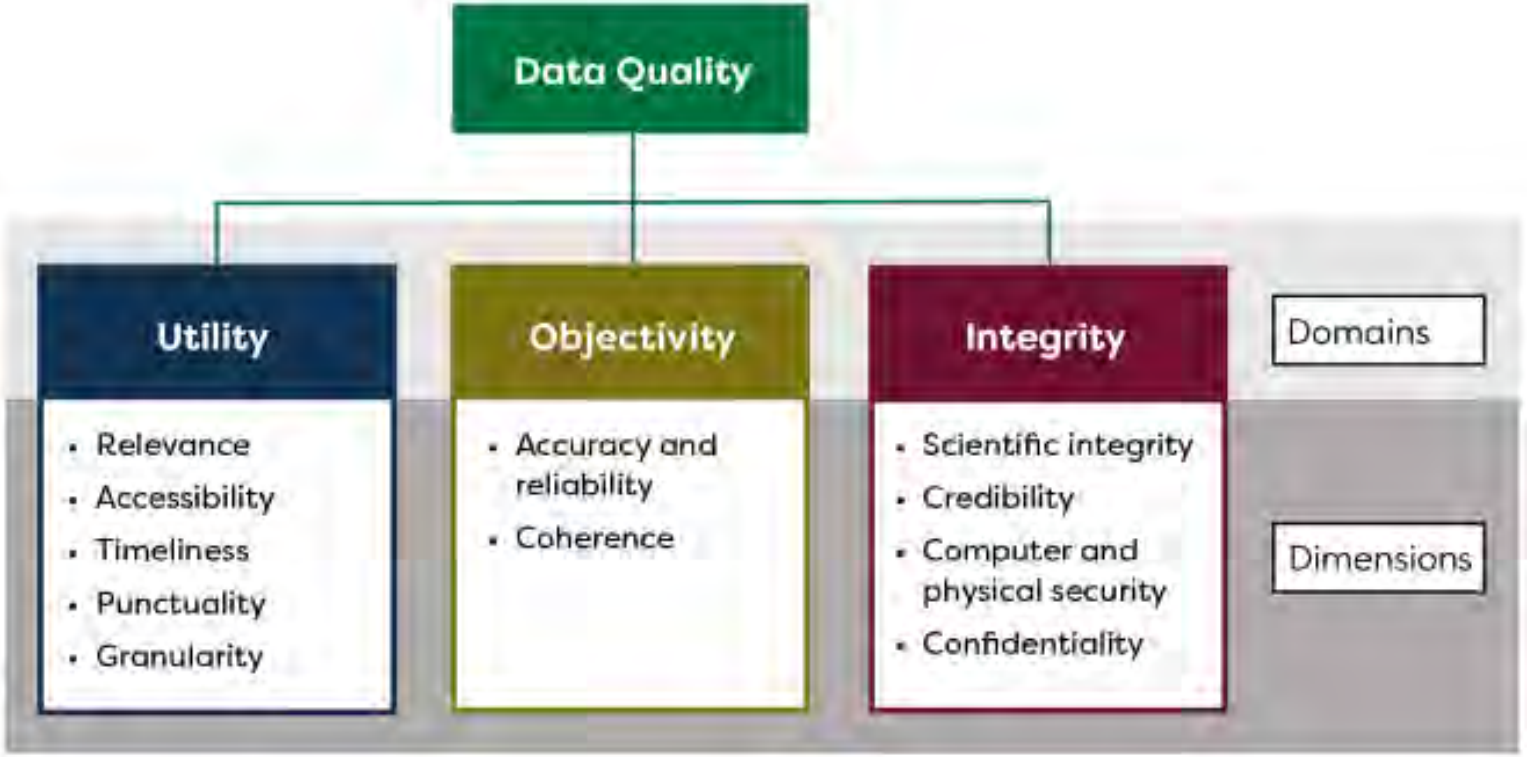
United States Census Bureau

Data Quality Today

- Traditionally data collection—including data discovery processes—is designed to minimize bias and maximize information content and verify quality of data after it is collected.
- This tradition is being challenged as we consider diverse uses, iterative processes, and building statistical products based on blended data through massive data repurposing.

2020 FCSM took on the challenge and proposed a data quality framework to support statistical products.

The FCSM Data Quality Framework



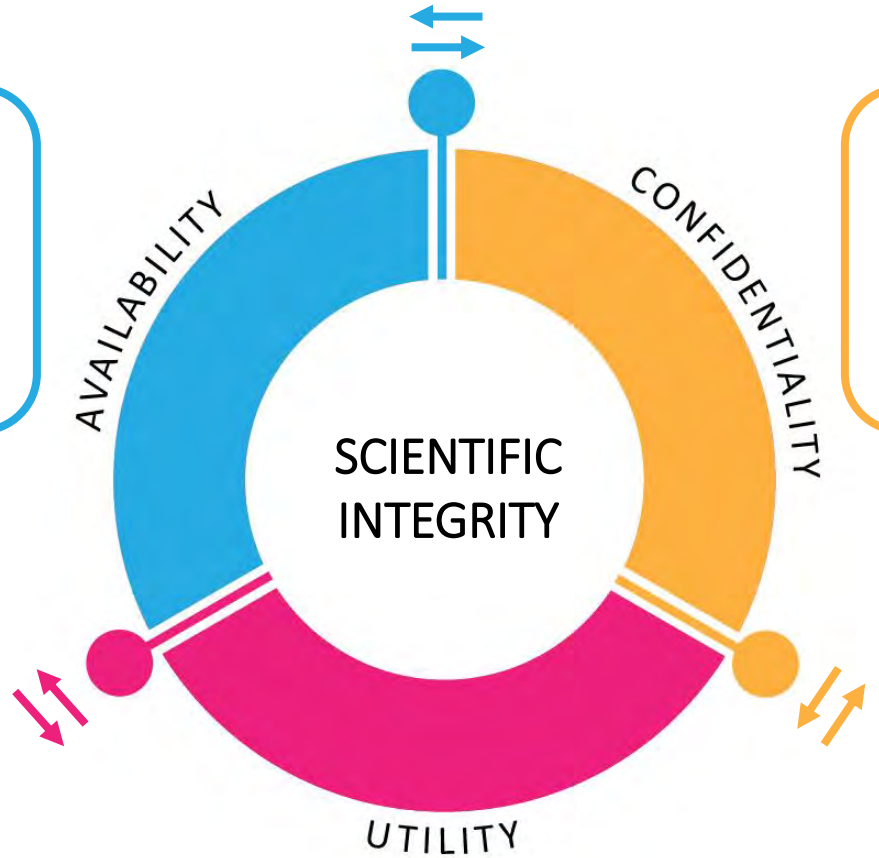
Building on the FCSM Framework to Operationalize a Statistical Quality Framework for the SPF Innovation Cycle

- Not a linear process.
- Simultaneously consider stakeholder engagement, data discovery, statistical methods, product design, while factoring in the needs of confidentiality.
- Scientific integrity—integral to the statistical quality framework—must be maintained throughout the SPF innovation cycle.

United States Census Bureau

The SPF Statistical Quality Framework

- Engagement
- Equity
- Granularity
- Timeliness
- Punctuality

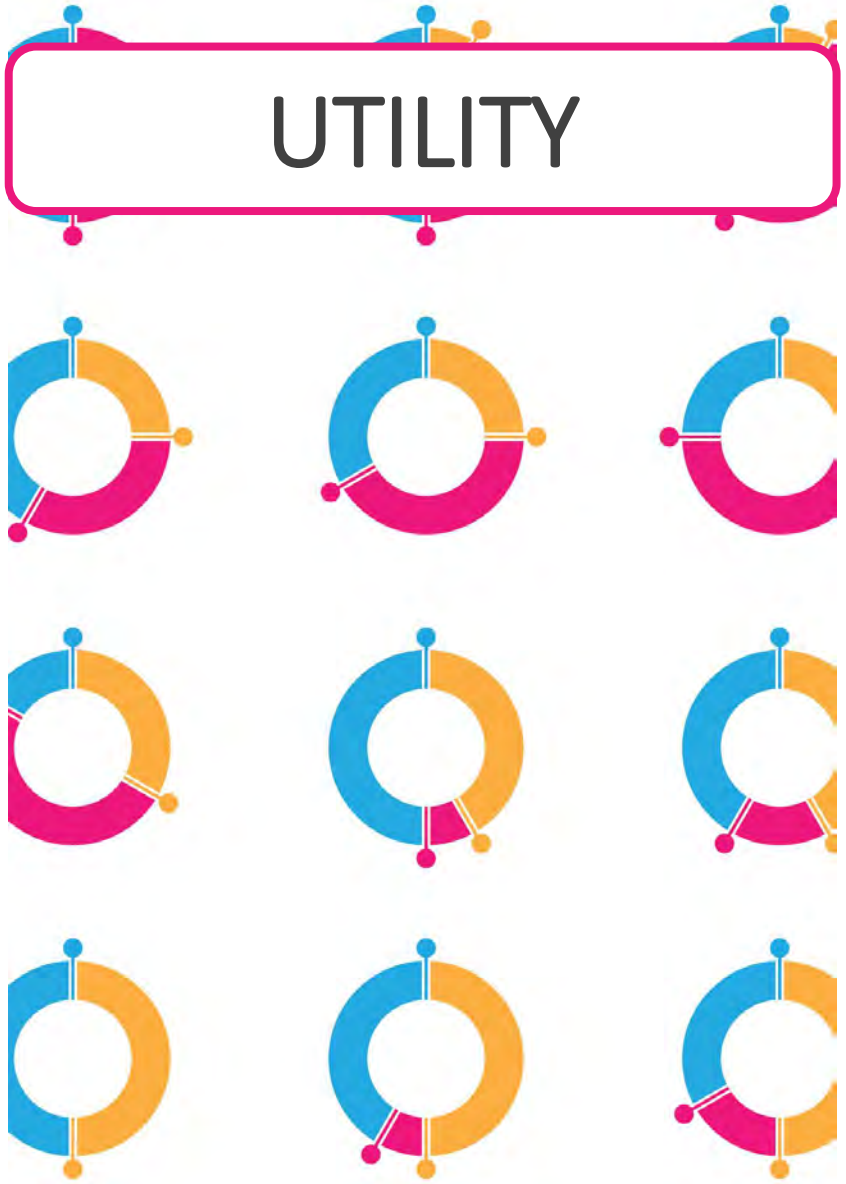


- Risk of Disclosure
- Risk of Harm

- Relevance
- Interpretability
- Accuracy/reliability
- Coherence
- Accessibility

United States Census Bureau

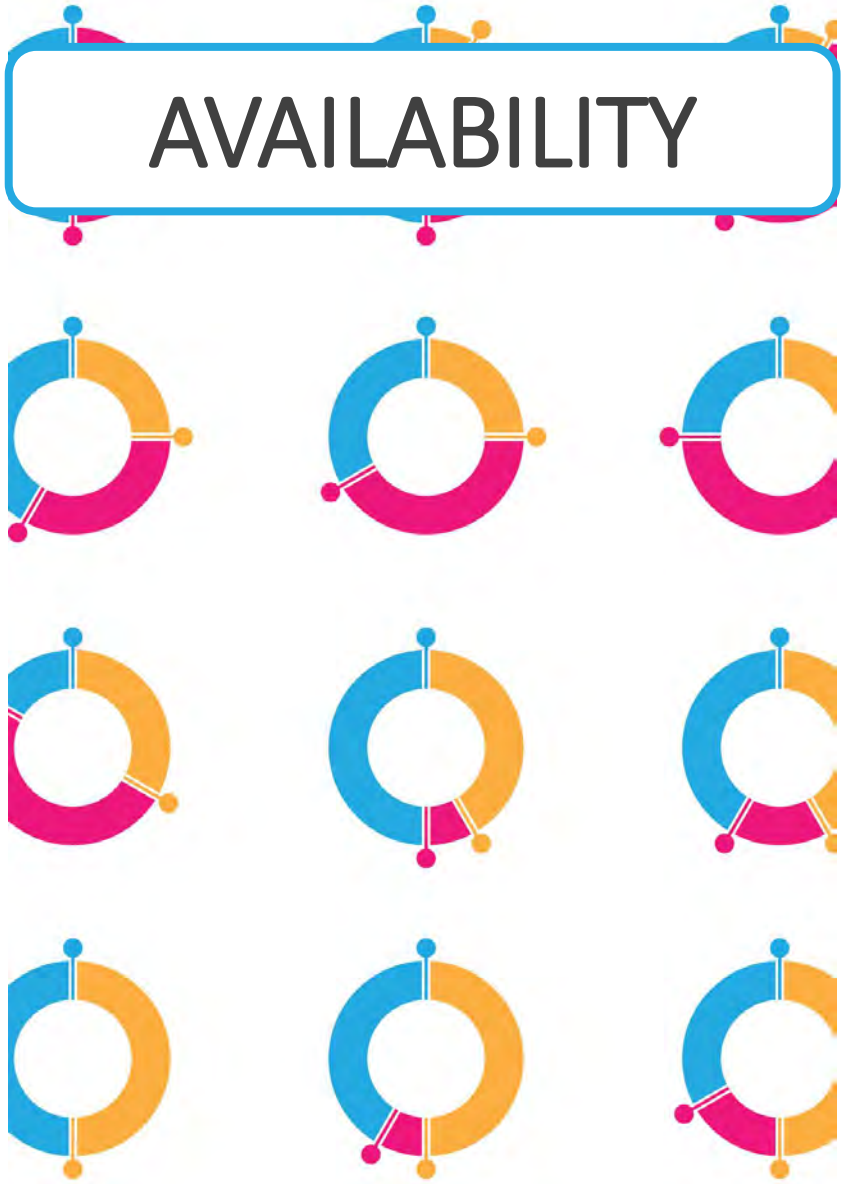
UTILITY



RELEVANCE	*What purposes and uses do Census statistics (i.e., statistical products) need to support?*
INTERPRETABILITY	*Will the statistics be readily understandable and interpretable by the intended data users?*
ACCURACY/ RELIABILITY	*Will the statistics be of sufficient accuracy and reliability to support valid statistical inference and decision-making for their intended uses?*
COHERENCE	*Will the statistics align sufficiently with other relevant statistical assets to permit broader purposes and uses?*
ACCESSIBILITY	*Can data users easily obtain the statistical products and documentation in forms and formats that are understandable to them?*

United States® Census Bureau

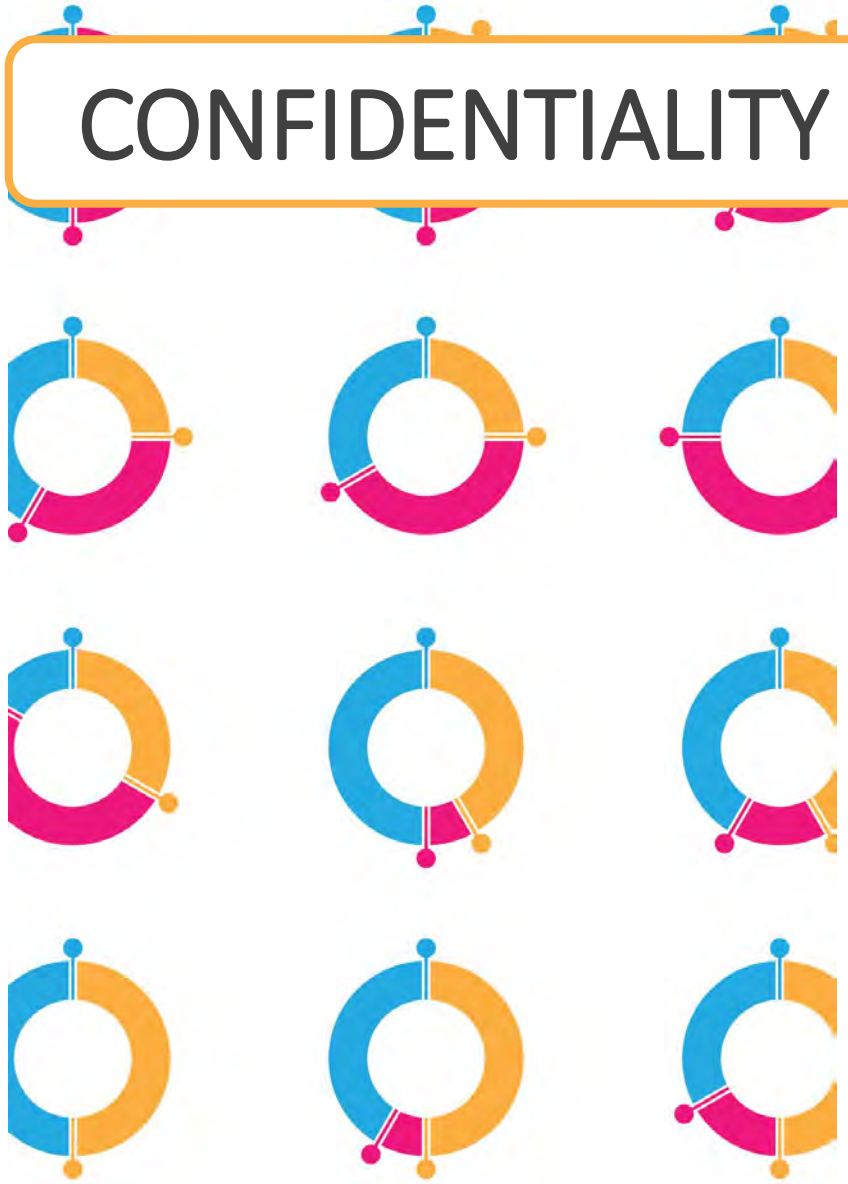
AVAILABILITY



ENGAGEMENT	*How can we iteratively engage with stakeholders to understand their purpose and use needs and verify the products developed meet their needs?*
GRANULARITY	*Are the statistics sufficiently disaggregated (by subgroups, geographies, and characteristics) to support purpose and use needs?*
EQUITY	*How well will the statistics support important uses by diverse groups of data users—whether demographic or various levels of data acumen?*
TIMELINESS	*How current do the statistics need to be to support their intended uses?*
PUNCTUALITY	*Are user expectations, through the development and dissemination of statistical products, being managed appropriately?*

United States® Census Bureau

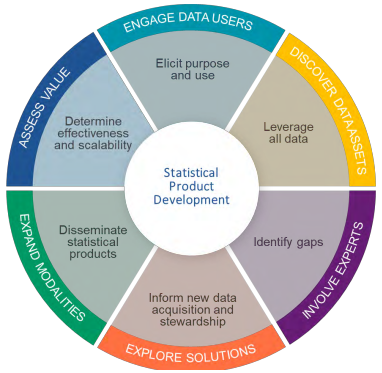
CONFIDENTIALITY



RISK OF DISCLOSURE	*What is the likelihood the statistics could lead to a disclosure that would reveal (or substantially improve inferences about) the identity or characteristics of a data subject?* *How reliable (certain) would those inference be?* *How strong is the confidentiality guarantee being given to data subjects?*
RISK OF HARM	*What is the risk of harm to a data subject were a disclosure to occur?* *How damaging might those harms be?* *What impact might this have on the agency's reputation?*

United States® Census Bureau

The Statistical Quality Framework—and all its dimensions—guides development



- Scientific integrity—integral to the statistical quality framework—undergirds the SPF innovation cycle.
- These 12 dimensions take on different nuances depending on where we are in the cycle.
- The ability to quantitatively vs. qualitatively assess development will also differ depending on where we are in the SPF innovation cycle.

UTILITY

- RELEVANCE
- INTERPRETABILITY
- ACCURACY/ RELIABILITY
- COHERENCE
- ACCESSIBILITY

AVAILABILITY

- ENGAGEMENT
- GRANULARITY
- EQUITY
- TIMELINESS
- PUNCTUALITY

CONFIDENTIALITY

- RISK OF DISCLOSURE
- RISK OF HARM

Exemplar: Grants.gov

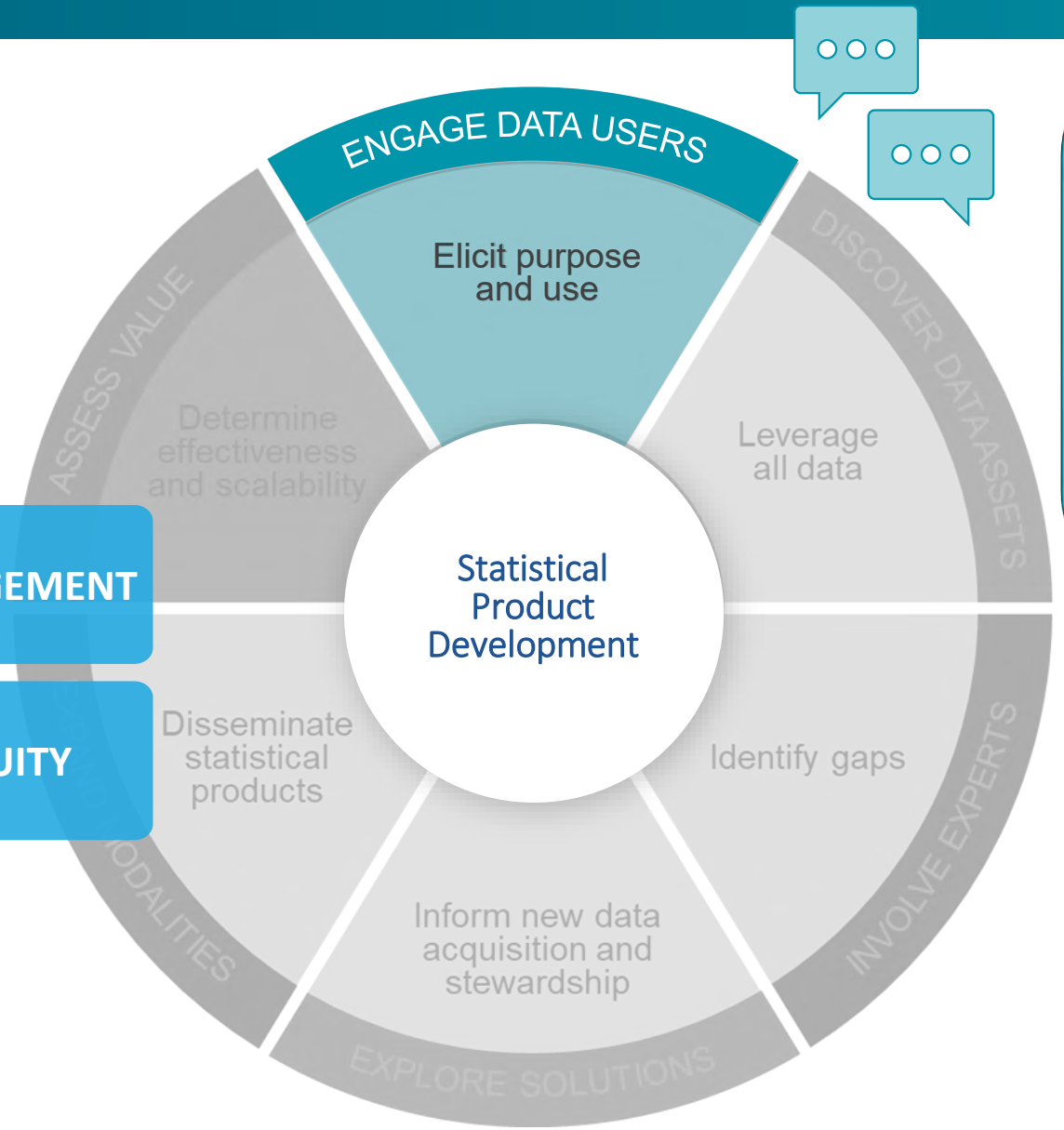
RELEVANCE

ENGAGEMENT

INTERPRETABILITY

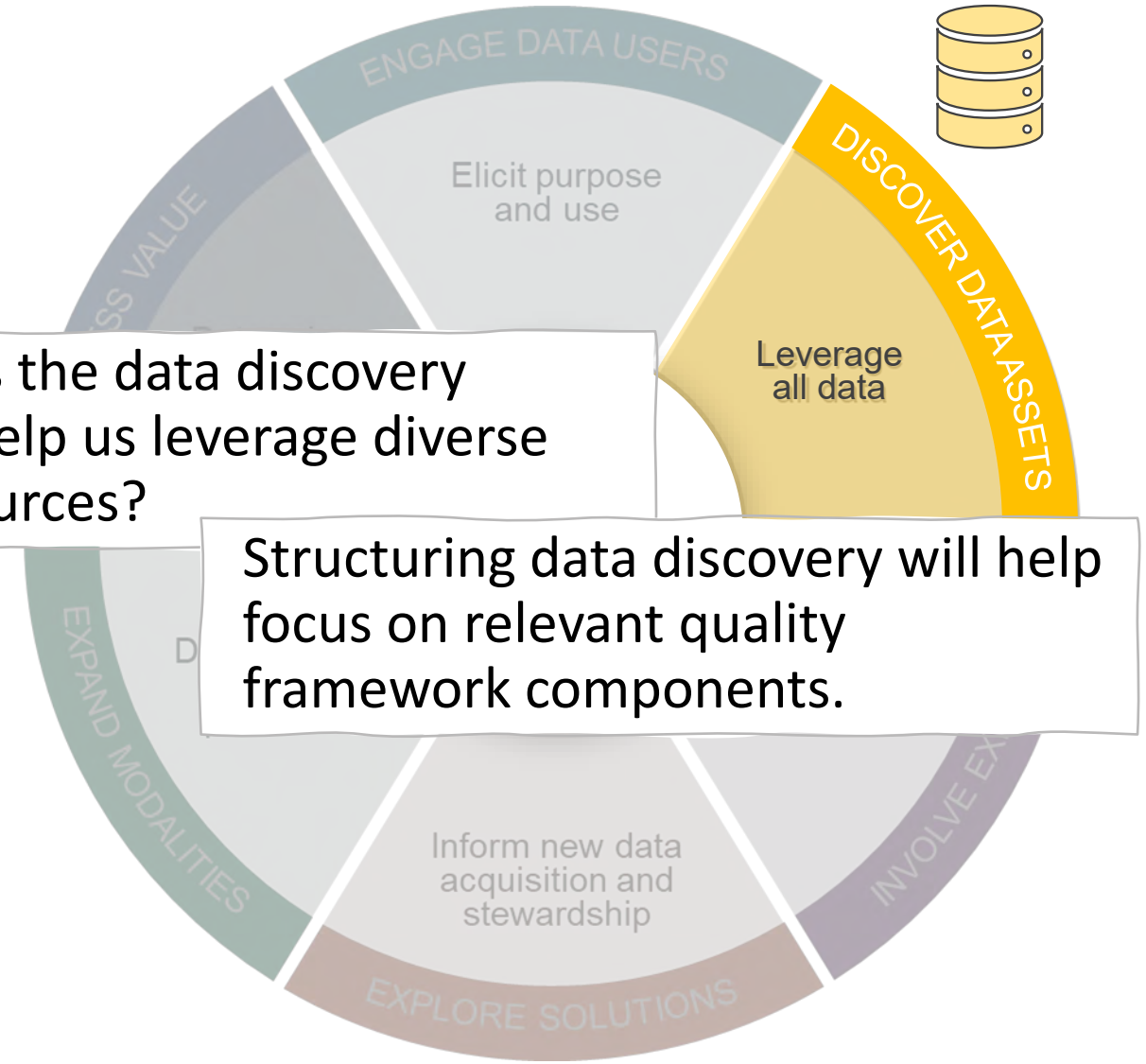
EQUITY

ACCESSIBILITY



We compiled information from listening sessions to create four personas that reflected stakeholders' needs and could be used to help support the goals of state and local governments, non-profits, and tribal organization members (SLaNT).

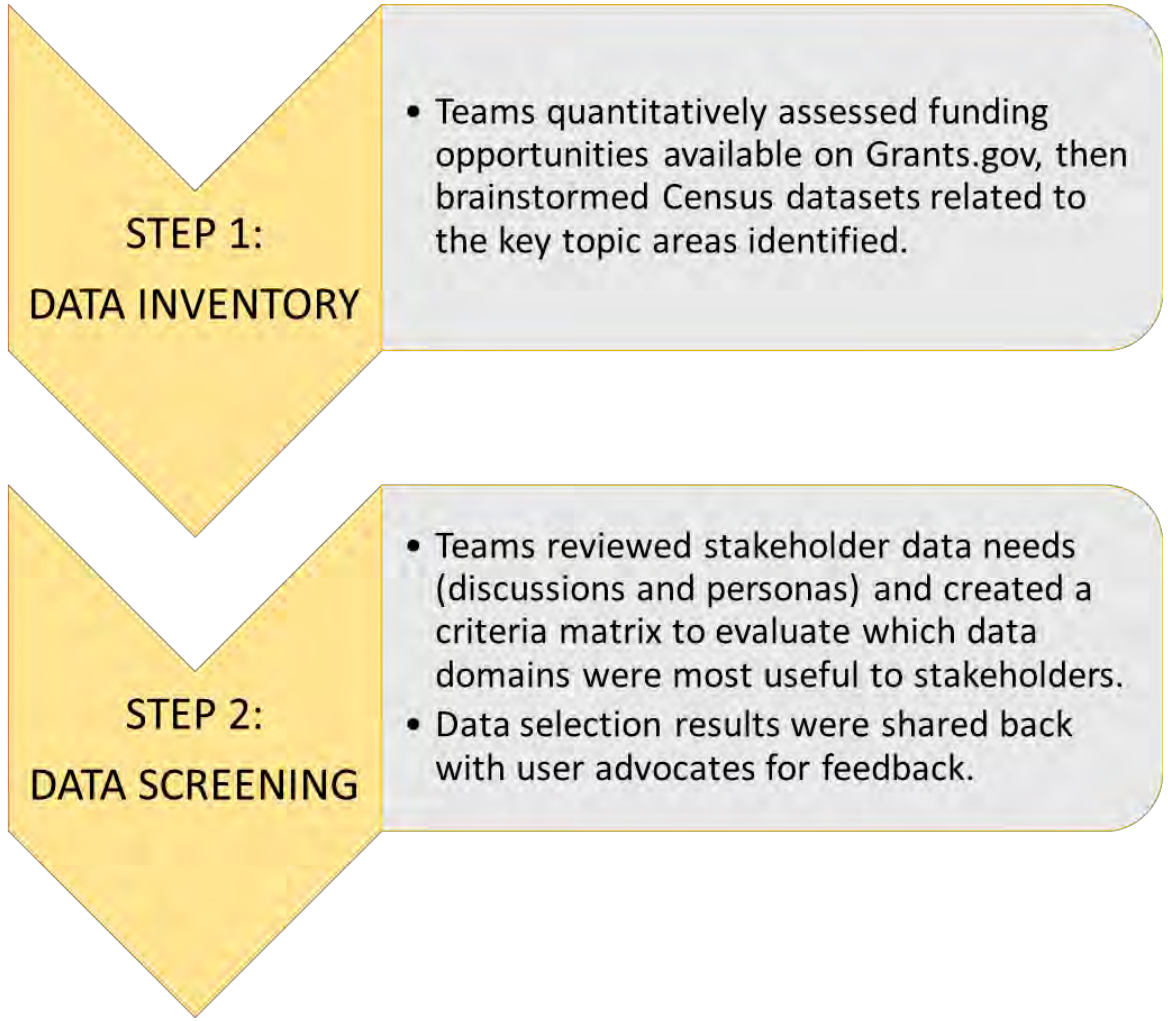
Supporting Grants.gov Applicants



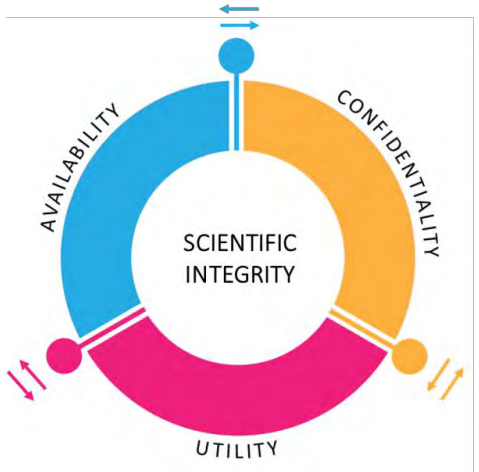
How does the data discovery process help us leverage diverse data resources?

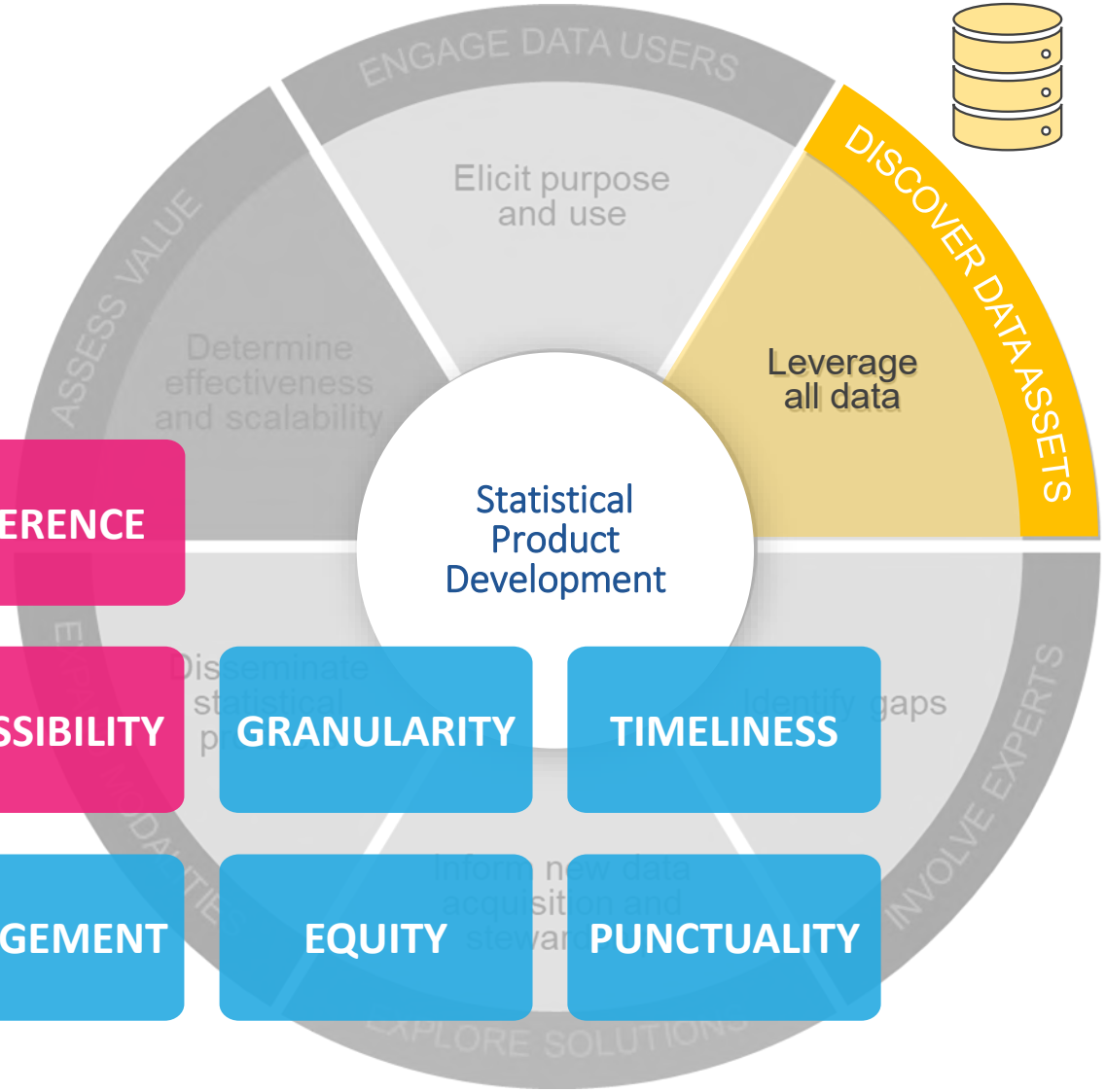
Structuring data discovery will help focus on relevant quality framework components.

How can we help organizations and applicants better utilize Census data when applying for funding via Grants.gov?



Supporting Grants.gov Applicants





RELEVANCE | **COHERENCE**

INTERPRETABILITY | **ACCESSIBILITY** | **GRANULARITY** | **TIMELINESS**

ACCURACY/ RELIABILITY | **ENGAGEMENT** | **EQUITY** | **PUNCTUALITY**

We analyzed all grants currently available on Grants.gov and identified the frequency and availability of funding opportunities by data domain.

Then, we identified potential Census data sources that support applications for the known funding opportunities.

United States Census Bureau

Data Screening Criteria Matrix

USER PERSONAS	INTERESTS	FUNDING CATEGORIES
Divya—a state public health official in Missouri	Education, Human Rights, Human Services, Transportation, Workforce Development	Education, Humanities, Transportation, Employment, Labor and Training, Income Security Social Services, Law Justice and Legal Services
Marcus—a city manager in Maine	Public Safety, Transportation, Parks and Recreation, Social Services, Job Creation, Business Growth, and Prosperity within his community	Transportation, Community Development, Humanities, Employment, Labor and Training, Business and Commerce
Sarah—a grant writer for a local community action agency in Pennsylvania	Poverty, Homelessness, Food Insecurity, Heating in Winter, and Other Social Challenges	Humanities, Food and Nutrition, Natural Resources, Income Security Social Services, Housing, Law Justice and Legal Services, Community Development
Charles—a community organizer in California	Community	Community Development

United States Census Bureau

Data Screening Criteria Matrix

USER PERSONAS	INTERESTS	FUNDING CATEGORIES
Divya—a state public health official in Missouri	Education, Human Rights, Human Services, Transportation, Workforce Development	Education, Humanities, Transportation, Employment, Labor and Training, Income Security Social Services, Law Justice and Legal Services
Marcus—a city manager in Maine	Public Safety, Transportation, Parks and Recreation, Social	Transportation, Community Development, Humanities, and
Sarah—a grant writer for a local community action agency in Pennsylvania	Poverty, Homelessness, Food Insecurity, Heating in Winter, and Other Social Challenges	Humanities, Food and Nutrition, Natural Resources, Income Security Social Services, Housing, Law Justice and Legal Services, Community Development
Charles—a community organizer in California	Community	Community Development

User interests influenced the data domain and measurement selection

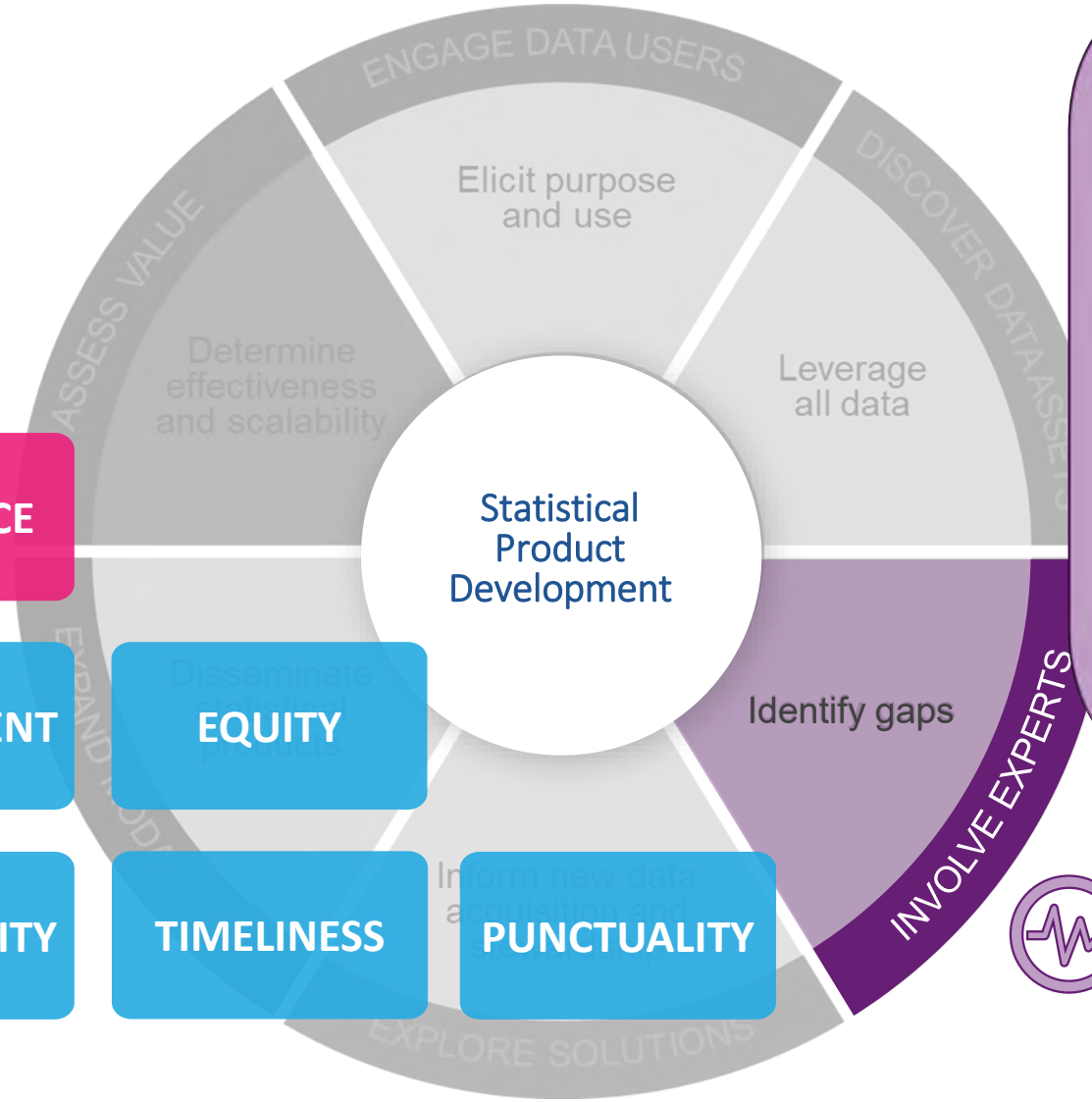
United States Census Bureau

Supporting Grants.gov Applicants

RELEVANCE | **COHERENCE**

INTERPRETABILITY | **ENGAGEMENT** | **EQUITY**

ACCURACY/ RELIABILITY | **GRANULARITY** | **TIMELINESS** | **PUNCTUALITY**

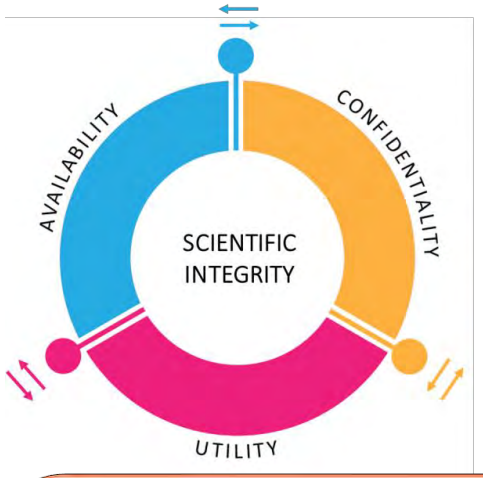


A criteria matrix was used to identify which data domains matter most to our stakeholders (e.g., health and nutrition, education, transportation, etc.).

Teams identified a list of 100 basic data elements that would help diverse SLaNT data users apply for grants.

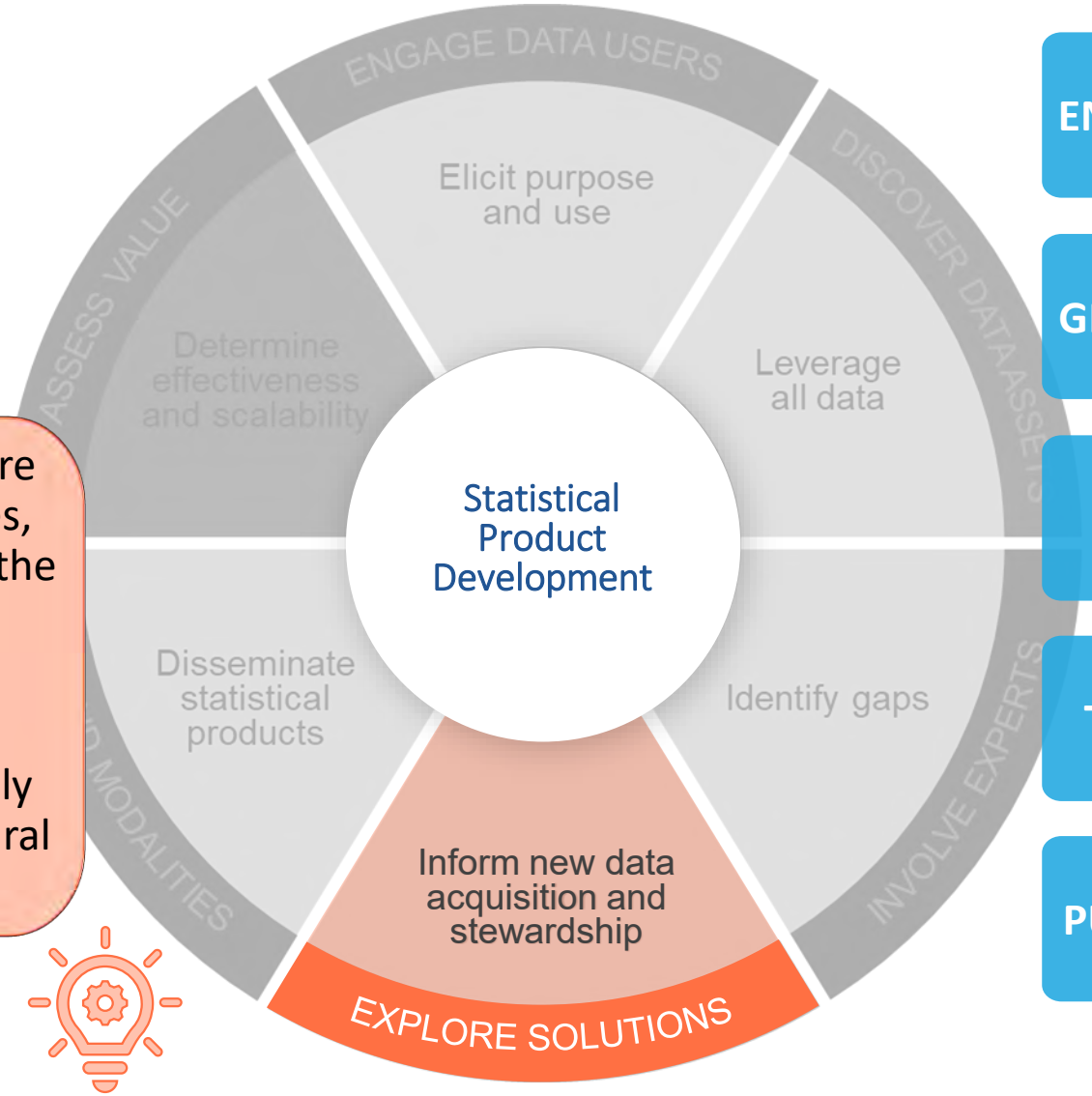
Results were shared back with internal experts and user advocates to identify what was missing.

Supporting Grants.gov Applicants



Granularity and timeliness were key concerns of user advocates, so data was selected to meet the screening criteria identified.

We identified non-Census, public data that might be needed to fill in gaps, especially around environment and natural resources.



ENGAGEMENT

RELEVANCE

GRANULARITY

INTERPRETABILITY

EQUITY

ACCURACY/ RELIABILITY

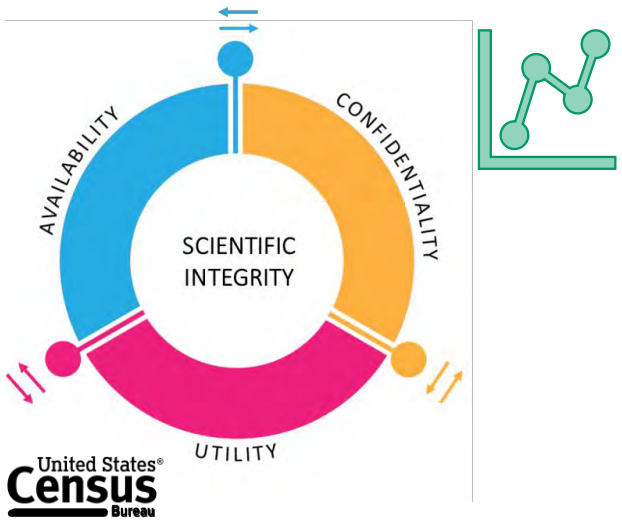
TIMELINESS

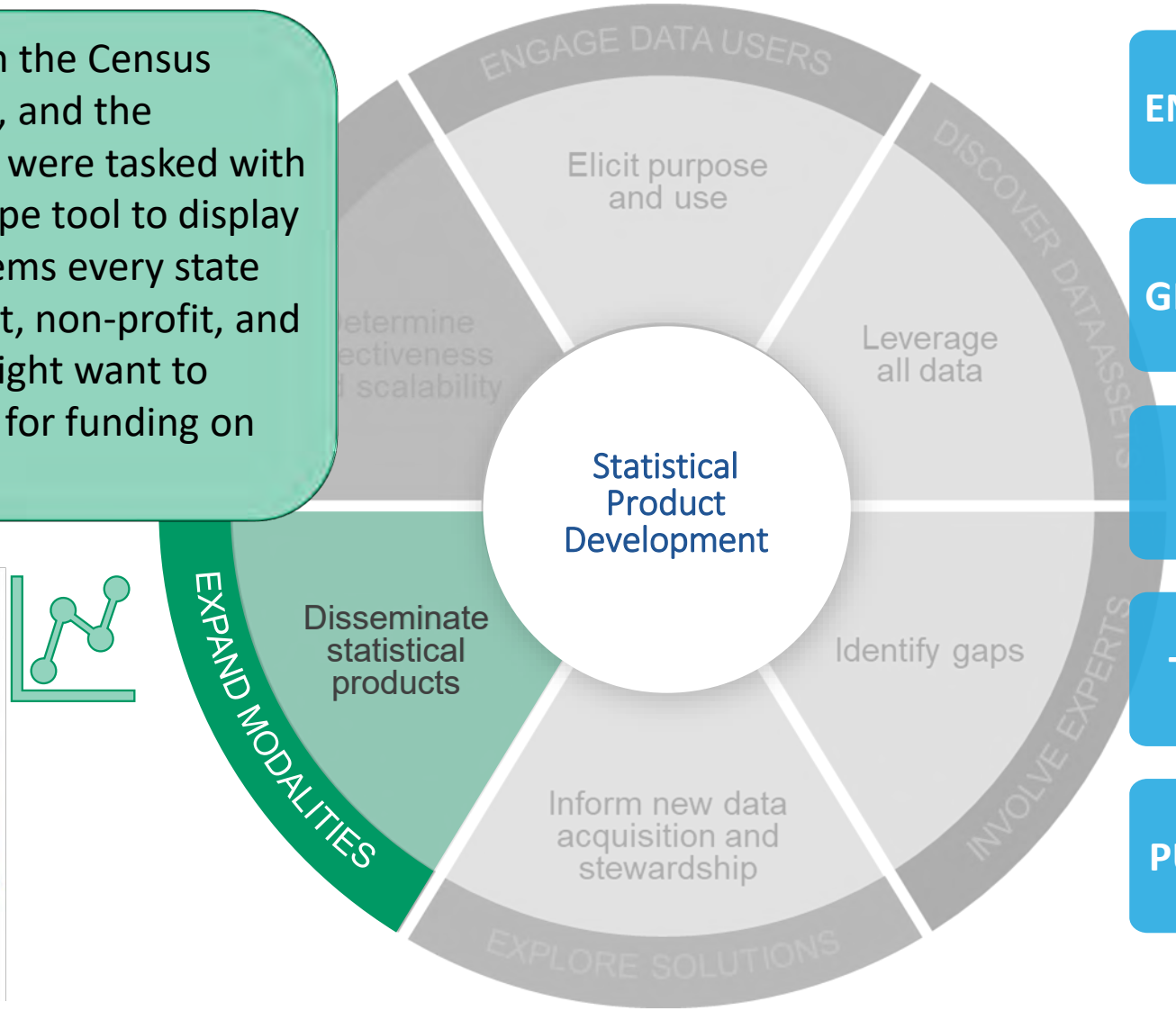
COHERENCE

PUNCTUALITY

Supporting Grants.gov Applicants

Capstone teams from the Census Bureau, Georgetown, and the University of Virginia were tasked with developing a prototype tool to display the 'Top 100' data items every state and local government, non-profit, and tribal organization might want to know when applying for funding on Grants.gov.





ENGAGEMENT | RELEVANCE

GRANULARITY | INTERPRETABILITY

EQUITY | ACCURACY/ RELIABILITY

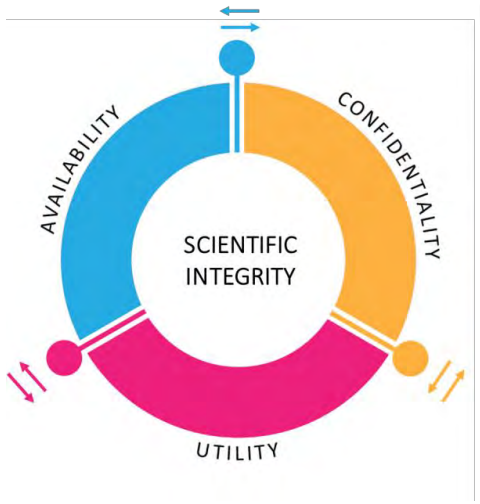
TIMELINESS | COHERENCE

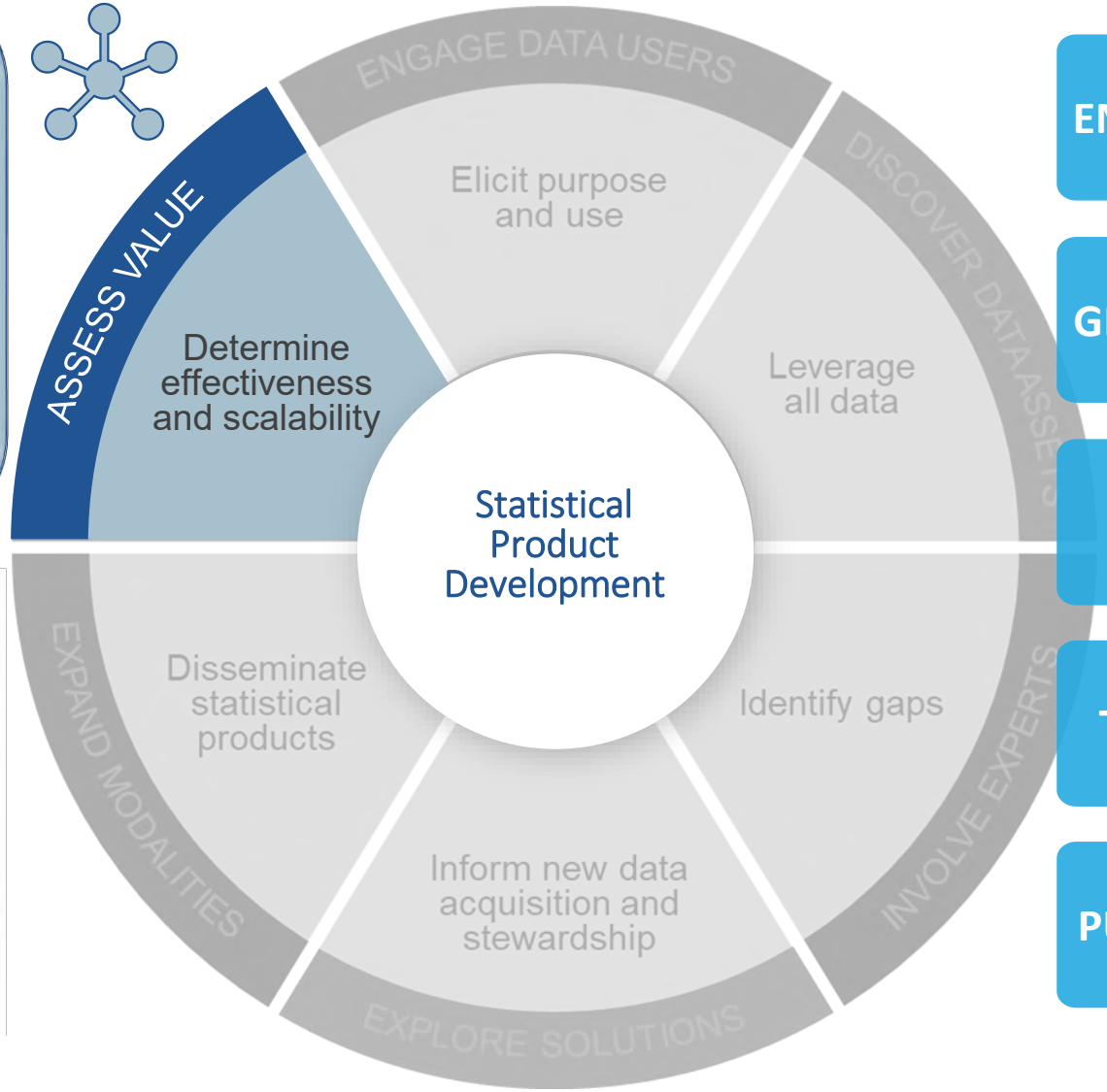
PUNCTUALITY | ACCESSIBILITY

United States® Census Bureau

Supporting Grants.gov Applicants

Plans are underway to share the 'Top 100' items and tool prototype with stakeholders and user advocates to identify what works and what doesn't before this tool goes through a second innovation cycle on its way to becoming a public facing statistical product.

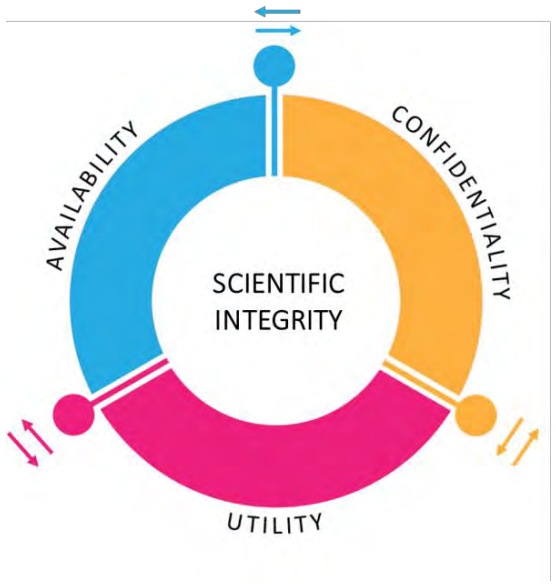




ENGAGEMENT
RELEVANCE
GRANULARITY
INTERPRETABILITY
EQUITY
ACCURACY/ RELIABILITY
TIMELINESS
COHERENCE
PUNCTUALITY
ACCESSIBILITY

United States Census Bureau

What about confidentiality?



RISK OF DISCLOSURE

RISK OF HARM

If new statistical products are a repurposing of public use, already disseminated statistics, there is no need to incorporate new decisions on confidentiality.

As we get user feedback and identify more gaps, new data acquisition choices and uses may introduce a new focus on confidentiality.

GDP-like Measure for Tribal Regions

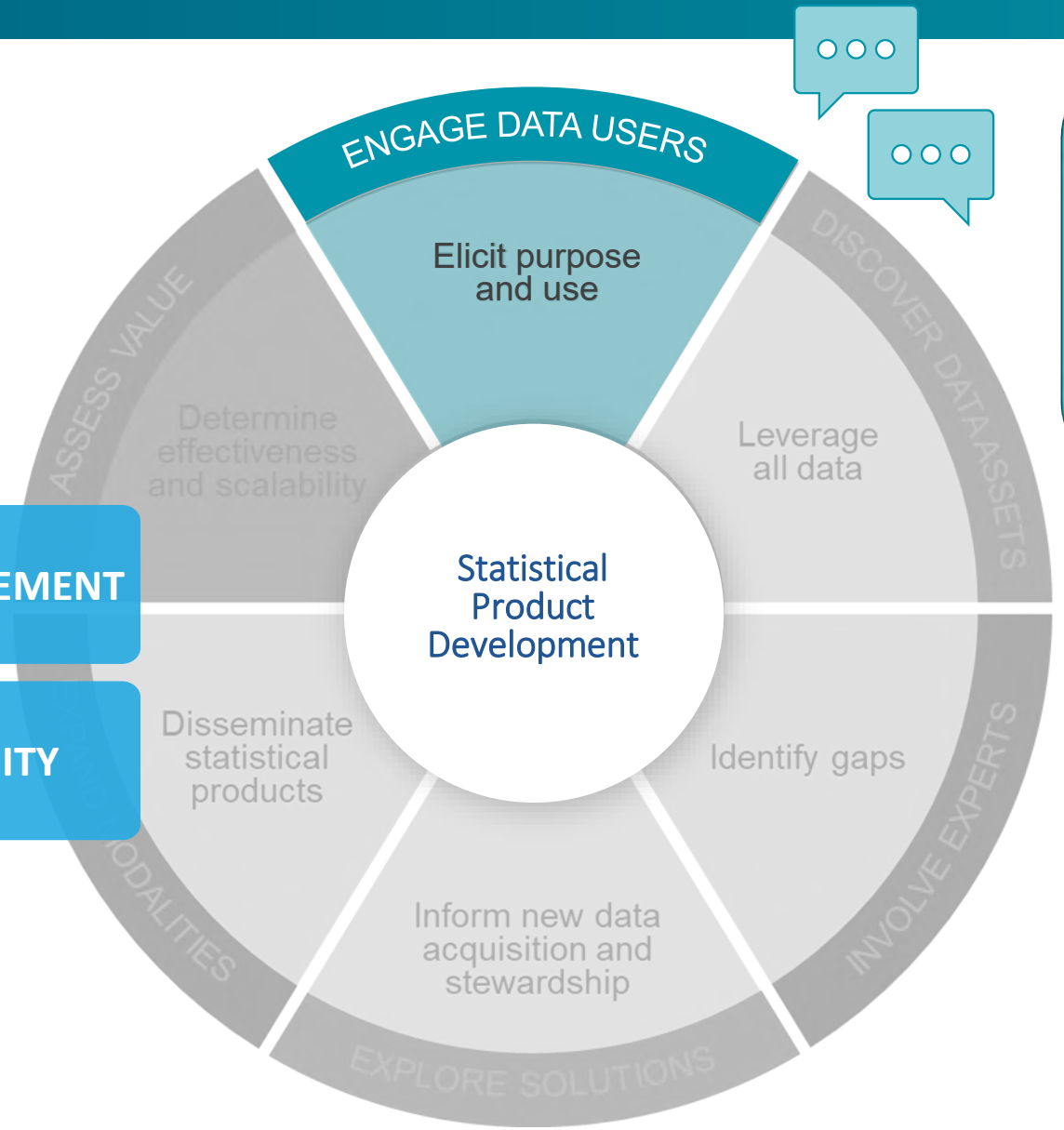
RELEVANCE

ENGAGEMENT

INTERPRETABILITY

EQUITY

ACCESSIBILITY

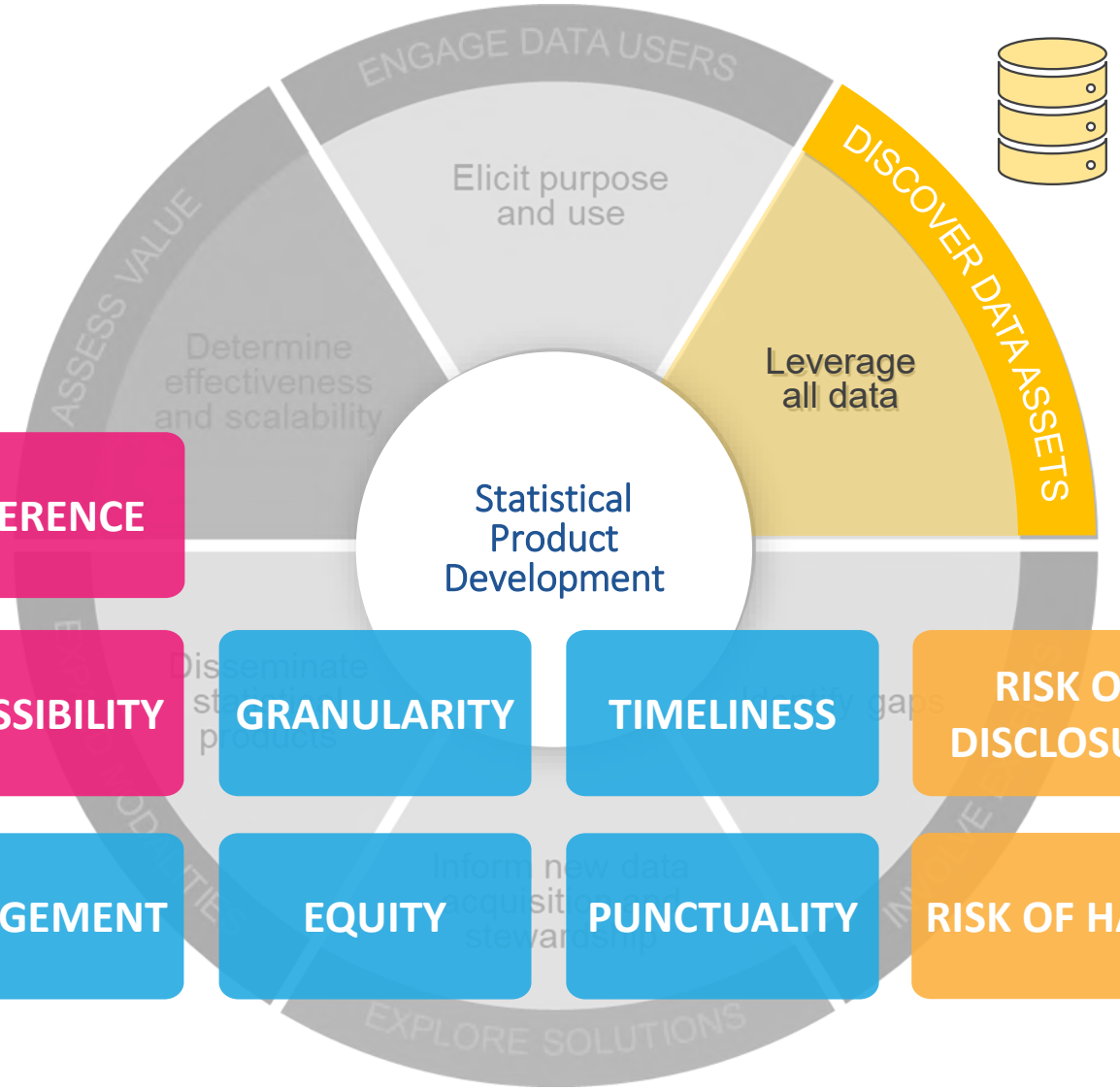


Data user engagements led to powerful purpose and use needs:

We need a GDP-like product to measure the economic health of tribal regions.

United States® Census Bureau

GDP-like Measure for Tribal Regions



RELEVANCE | **COHERENCE**

INTERPRETABILITY | **ACCESSIBILITY** | **GRANULARITY** | **TIMELINESS** | **RISK OF DISCLOSURE**

ACCURACY/ RELIABILITY | **ENGAGEMENT** | **EQUITY** | **PUNCTUALITY** | **RISK OF HARM**

The Census Bureau collects vast amounts of economic data and then provides the estimates to the Bureau of Economic Analysis, where official GDP measures are produced.

United States Census Bureau

GDP-like Measure for Tribal Regions

RELEVANCE

COHERENCE

EQUITY

TIMELINESS

PUNCTUALITY

INTERPRETABILITY

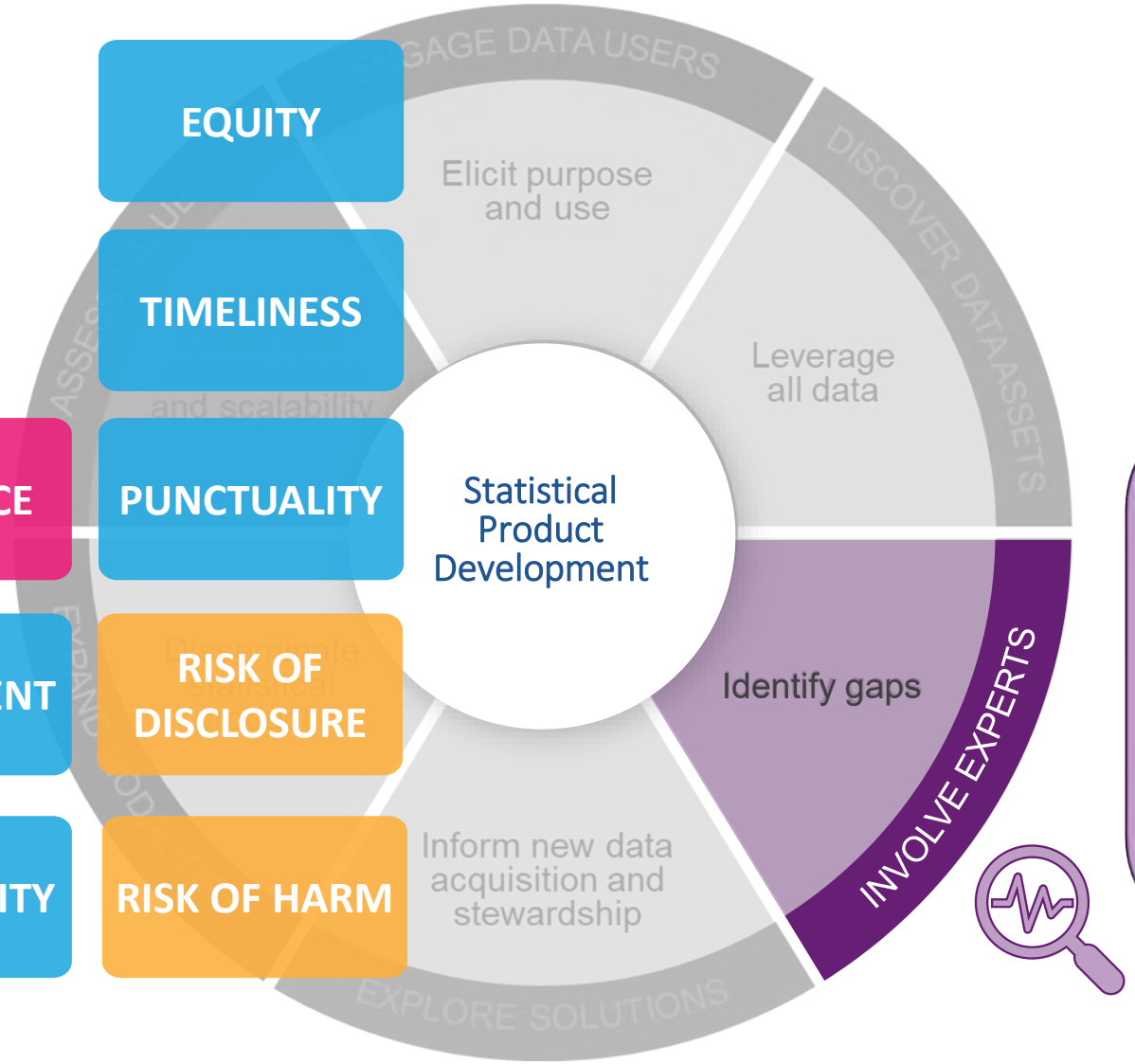
ENGAGEMENT

RISK OF DISCLOSURE

ACCURACY/ RELIABILITY

GRANULARITY

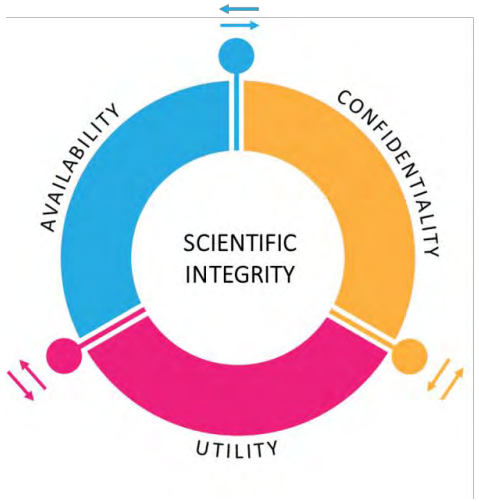
RISK OF HARM



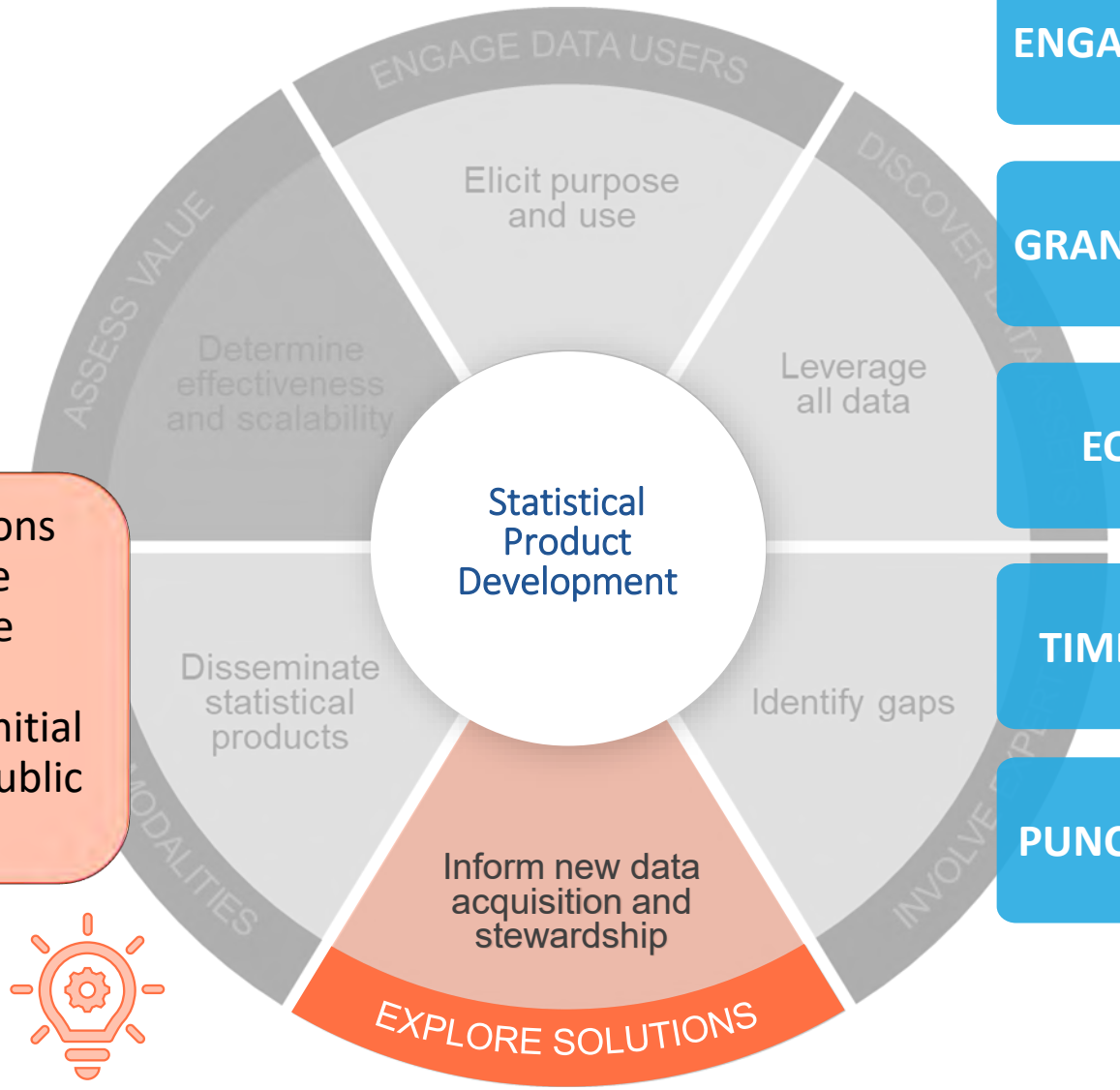
The GDP formula includes government spending. The Census Bureau collects this information as part of its Census of Governments. However, tribal nations are not included in that data collection.

United States Census Bureau

GDP-like Measure for Tribal Regions



Currently exploring options for calculating a GDP-like measure but without the component of tribal government spending. Initial statistics are based on public use data sources.



ENGAGEMENT

RELEVANCE

GRANULARITY

INTERPRETABILITY

EQUITY

ACCURACY/ RELIABILITY

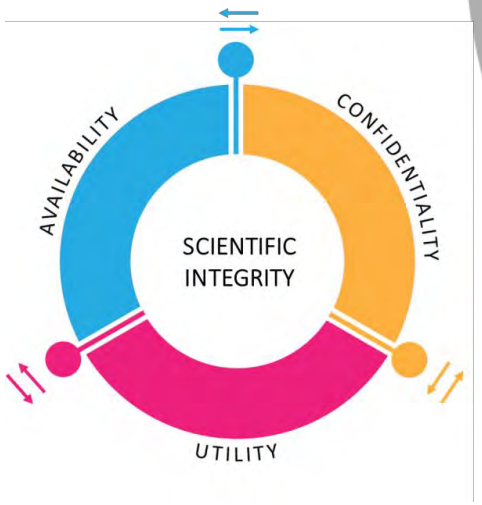
TIMELINESS

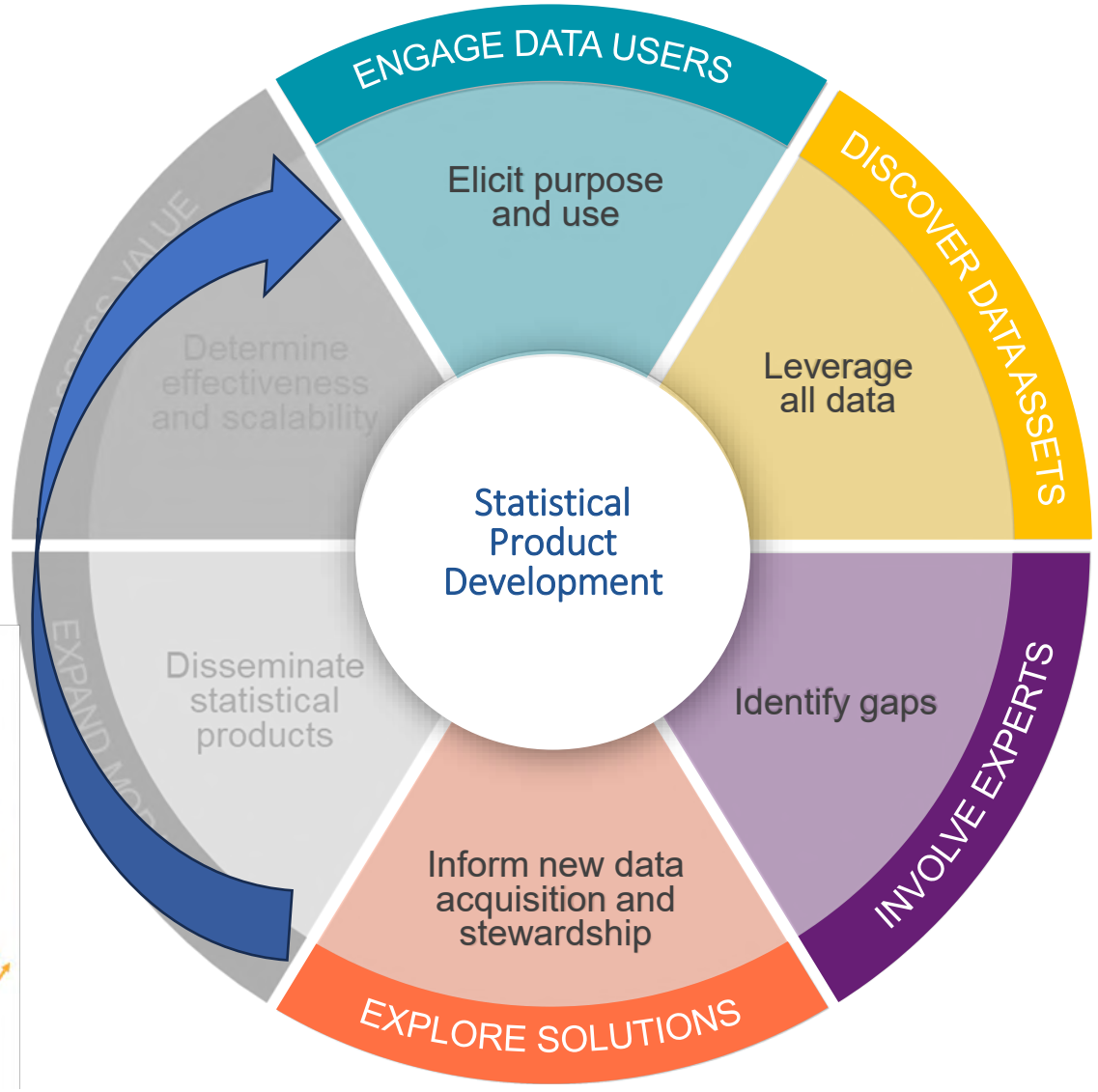
COHERENCE

PUNCTUALITY

GDP-like Measure for Tribal Regions

Using early estimates to ground more engagements and discussions around purpose and use needs. These will inform more data discovery of confidential data assets and the refinement of new tribal GDP-like statistical products.





ENGAGEMENT	RELEVANCE
GRANULARITY	INTERPRETABILITY
EQUITY	ACCURACY/ RELIABILITY
TIMELINESS	COHERENCE
PUNCTUALITY	ACCESSIBILITY
RISK OF HARM	RISK OF DISCLOSURE

United States Census Bureau


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

United States® Census Bureau