

The U.S. Federal Statistical System

ACDEB – 1/22/2021



Agenda

An Introduction To The Federal Statistical System

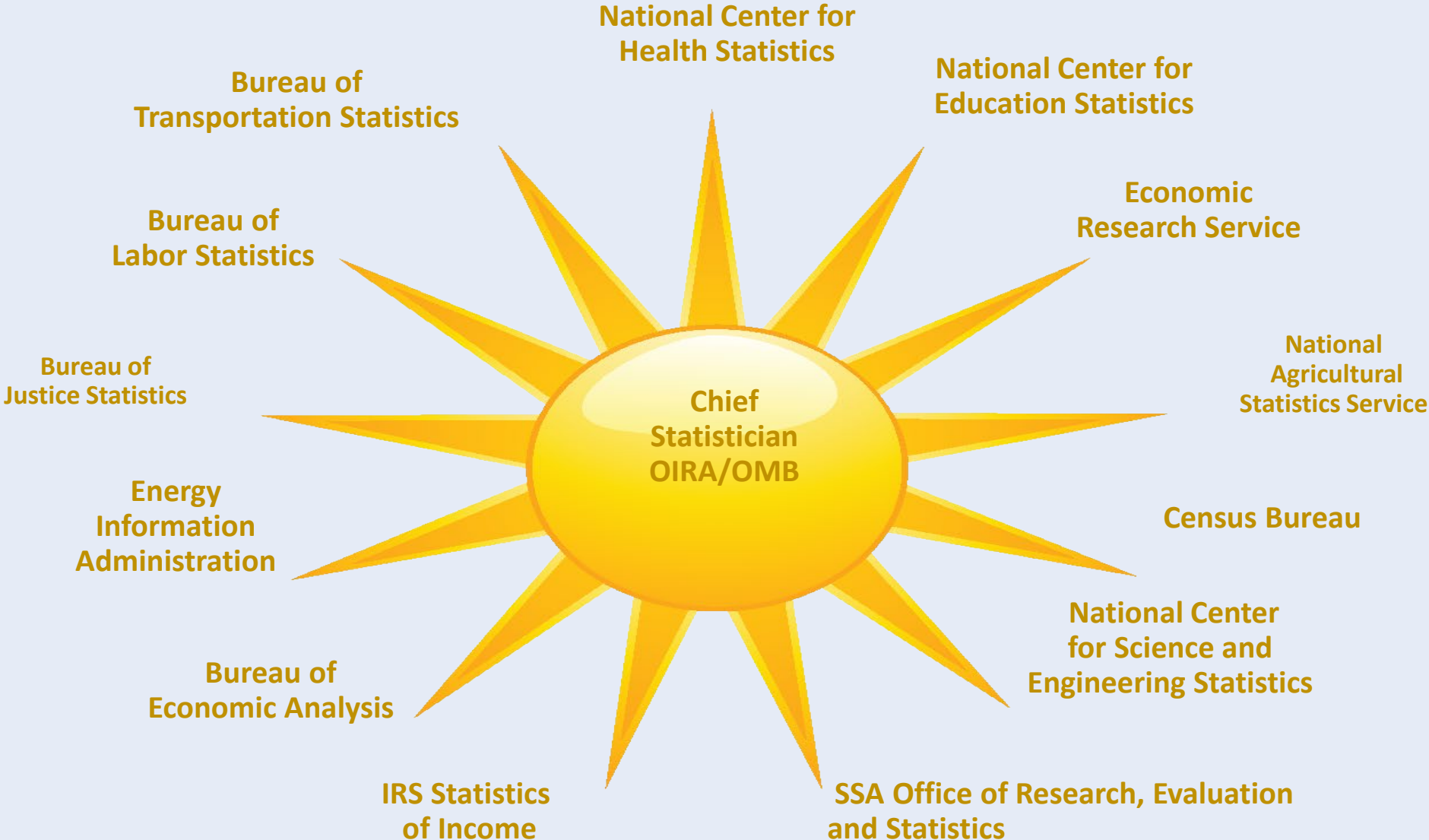
- Who We Are—Brian Moyer
- What We Do—Barry Johnson
- Where We Are Going—Emilda Rivers



Who we are

Trusted sources of data that the public, businesses, and governments need to make informed decisions. Timely, accurate, objective, and relevant statistical data are the foundation of evidence-based decision-making.

Principal Statistical Agencies



A Decentralized Structure

- 13 principal statistical agencies
- 90 plus additional agencies and units engaging in statistical activities
- Statistical and Science Policy Office, Office of Management and Budget

What is a Statistical Agency?

The term “statistical agency or unit” means an agency or organizational unit of the executive branch whose activities are predominantly the collection, compilation, processing, or analysis of information for statistical purposes...

Interagency Council on Statistical Policy (ICSP)

- Started in 1987 to improve communication among the statistical agencies
- Evidence Act expanded membership to include all Statistical Officials
- Provides a forum for inter-agency information sharing and collaboration on shared goals and challenges
- Sets Strategic goals for [modernizing](#) the agency practices and products and advances those goals by collaborating on strategic initiatives
 - Example: Goal to build a strong workforce has resulted in an annual, interagency mentoring program, shared training opportunities, innovation sharing events (Big Data and Data Linkage Days)
- Plays a leading role in implementing the Evidence Act and advancing the Federal Data Strategy


OMB Statistical Policy and Coordination Functions

- Ensure that agencies' budget proposals are consistent with Statistical System priorities
- Develop and implement government-wide policies, principles, standards, and guidelines
- Evaluate statistical program performance and agency compliance with policies, principles, standards, and guidelines
- Approve all agency information collections

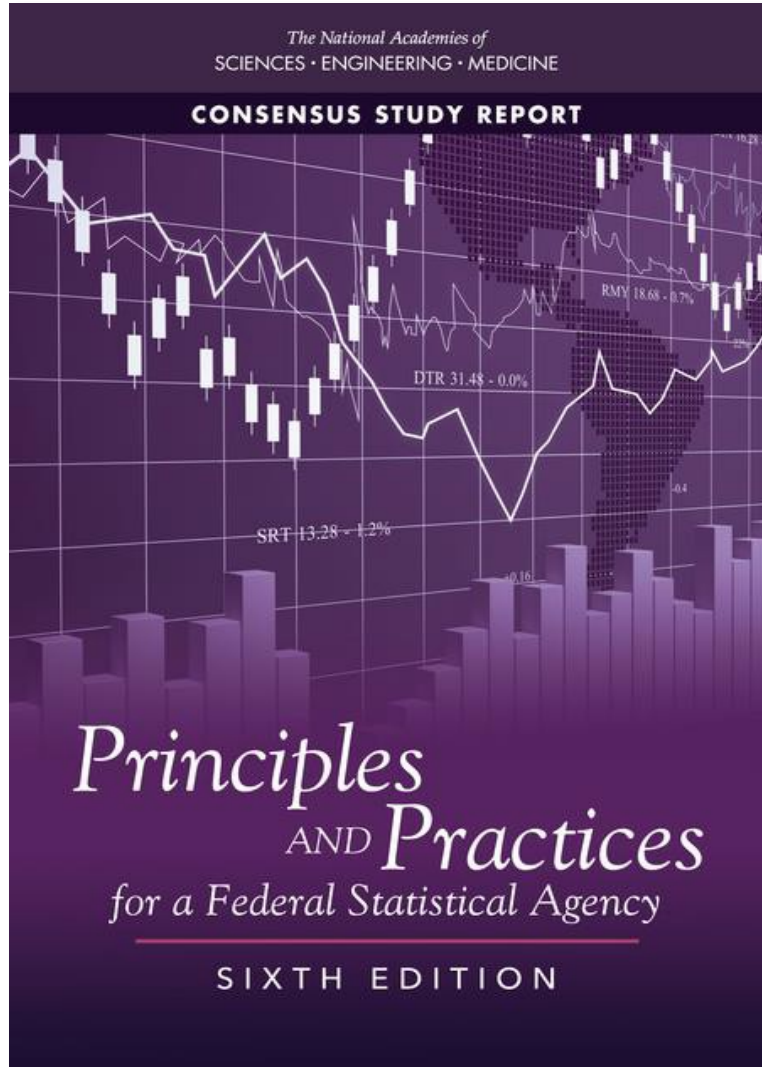
OMB Statistical Policy and Coordination Functions

- Promote the sharing of statistical information consistent with privacy rights and confidentiality pledges
- Coordinate participation in international statistical activities, including the development of comparable statistics
- Lead and work with interagency councils to facilitate statistical functions and activities
- Provide opportunities for training in statistical policy functions to employees of the Federal Government

Statistical Standards and Guidelines

- OMB Statistical Directives
 - Directive 1: *Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units*
 - Directive 2: *Standards and Guidelines for Statistical Surveys*
 - Directive 3: *Compilation, Release, and Evaluation of Principal Federal Economic Indicators*
 - Directive 4: *Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies*
 - *Classification standards (e.g., NAICS, race and ethnicity)*
 - Confidential Information Protection and Statistical Efficiency Act (CIPSEA) and Implementation Guidance
 - Evidence Act and Federal Data Strategy
- 

Other Sources of Statistical Standards

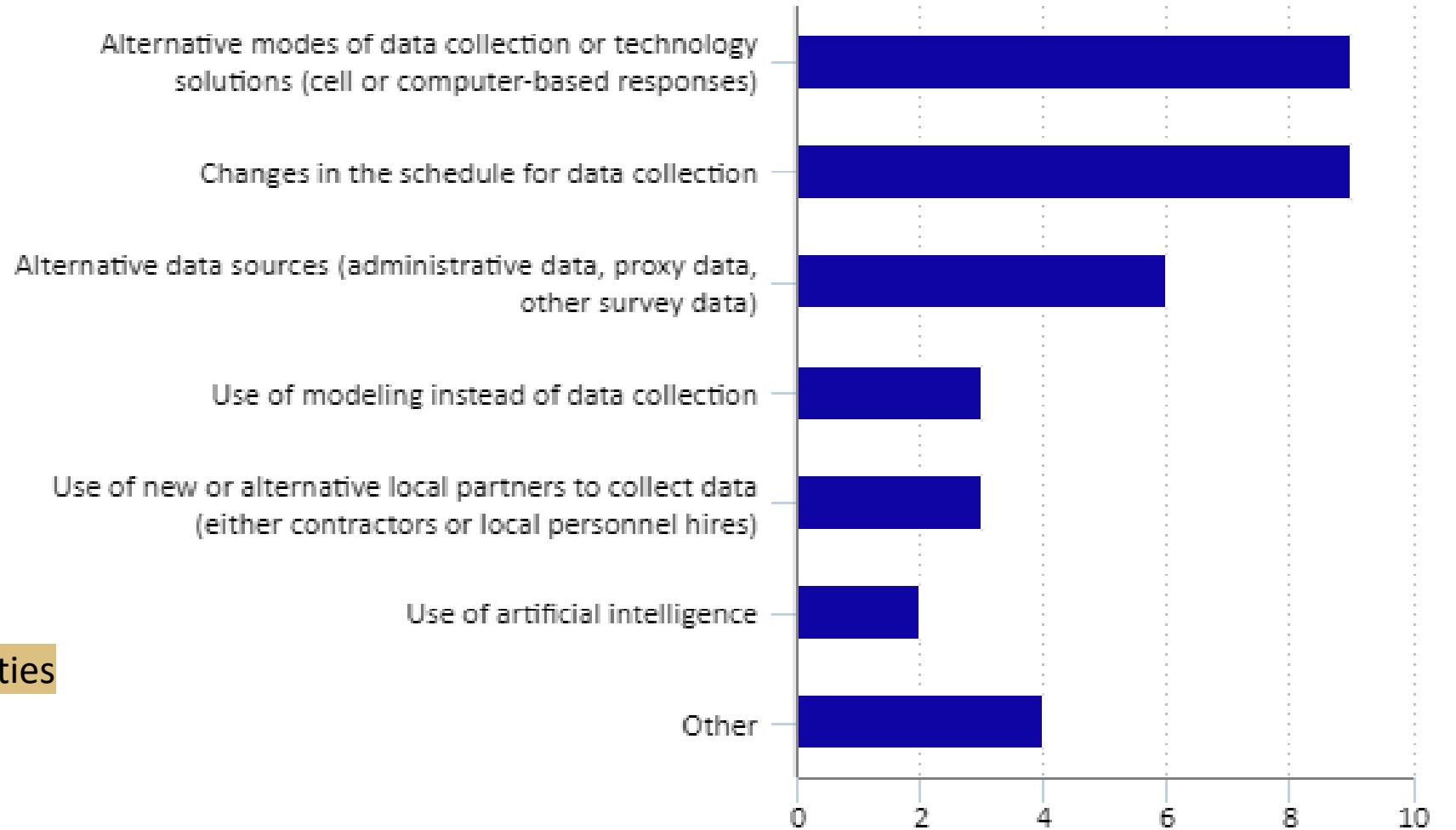


Produced by the National Academies
of Sciences, Engineering and
Medicine's Committee on National
Statistics

Cited in OMB Statistical Directive #1

Innovating Amid COVID-19

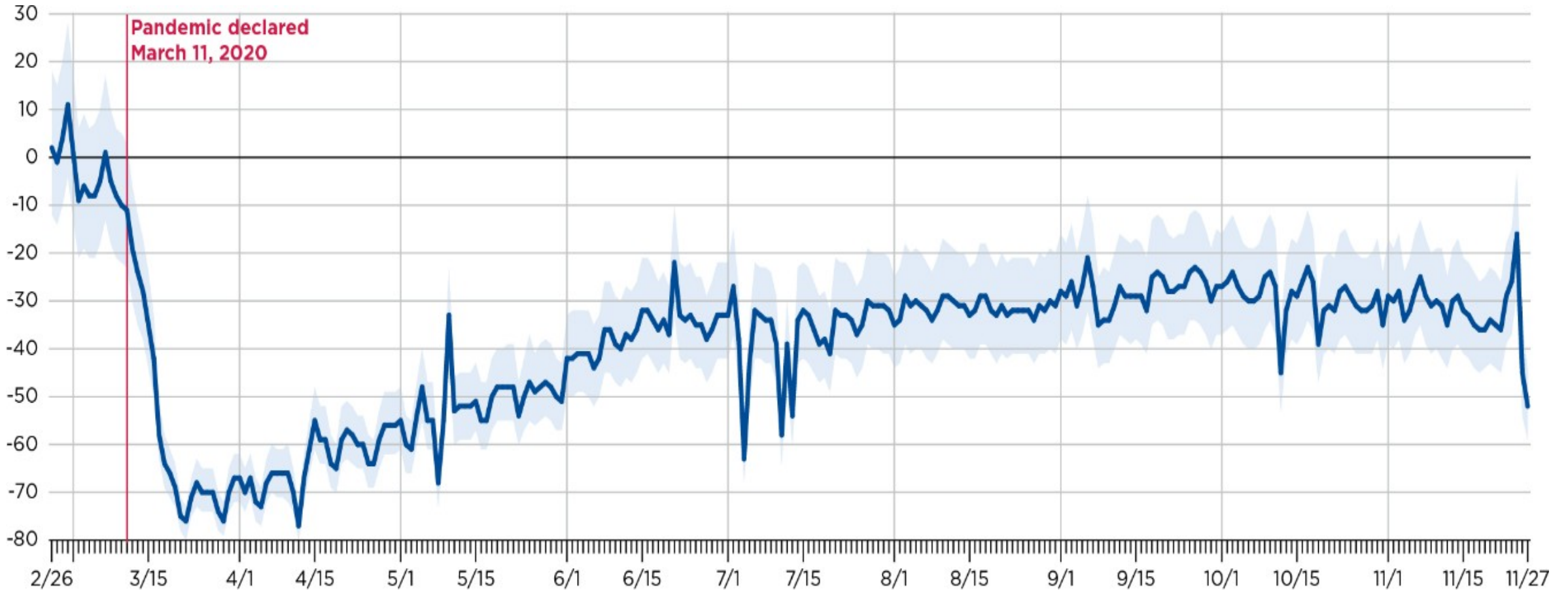
Did your agency use any of the following in place of planned or existing data collection methods during the pandemic?



Innovating Around High Frequency And Blended Data

Innovation Spotlight: Bureau of Economic Analysis (BEA) Near Real-Time Consumer Spending Statistics

Daily Spending on Food Services and Drinking Places [Percent]





The Household Pulse Survey – collaborative innovation

***NEW: Statistical Agency Changes in Response to the COVID-19 Pandemic**

// Census.gov > Data > Experimental Data Products > Household Pulse Survey

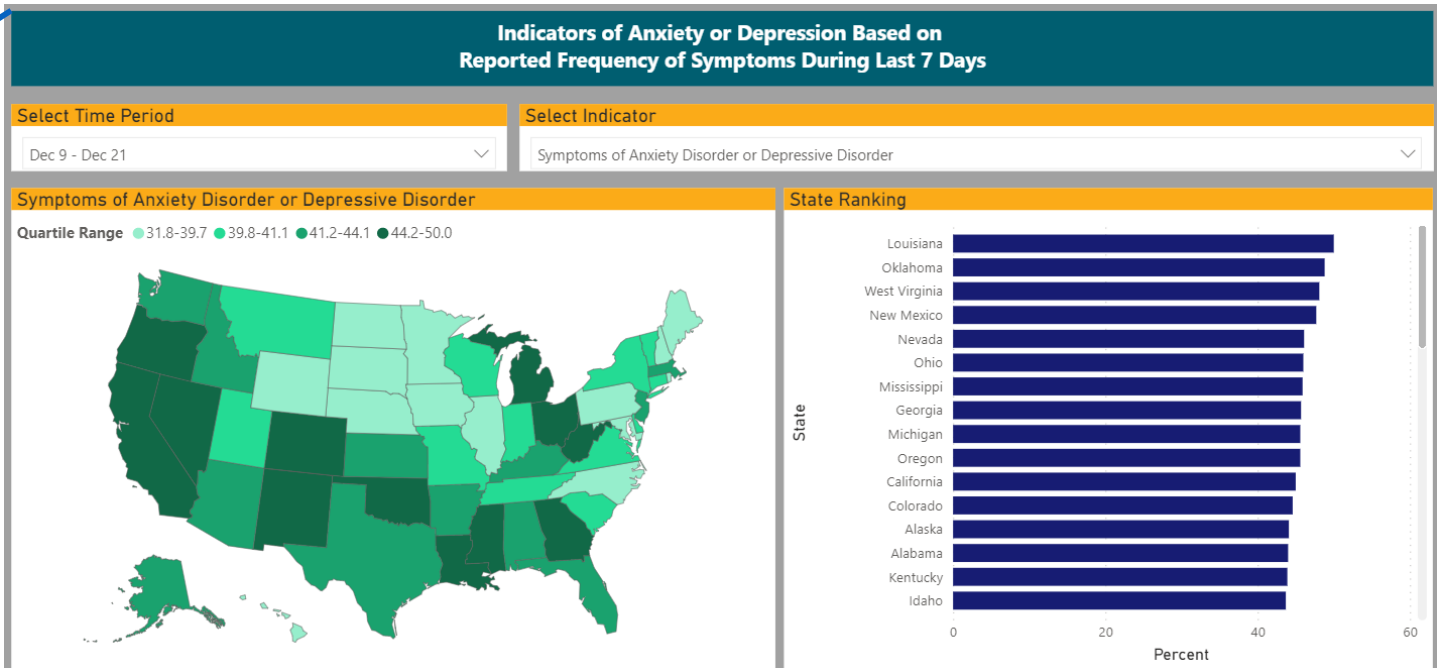
Measuring Household Experiences during the Coronavirus Pandemic



Household Pulse Survey – Phase 3
(October 28, 2020 – March 1, 2021)

What is the Household Pulse Survey?

The U.S. Census Bureau, in collaboration with multiple federal agencies, is in a unique position to produce data on the social and economic effects of coronavirus on American households. The Household Pulse Survey is designed to deploy quickly and efficiently, collecting data to measure household experiences during the coronavirus pandemic. Data will be disseminated in near real-time to inform federal and state response and recovery planning.





The Evidence Act's 3 Big Ideas

- Evidence-Based Policymaking Requires Systematic Planning
- Effective Federal Data Asset Use Requires High-Quality Data Governance (including Open Data Planning)
- Safely Expanding Access to Protected Data for Evidence-building

Title III & the Role of Statistical Agencies



Strengthens technical
autonomy of
agencies



Provides new
authority for
agencies to acquire
data for the
purpose of
evidence building

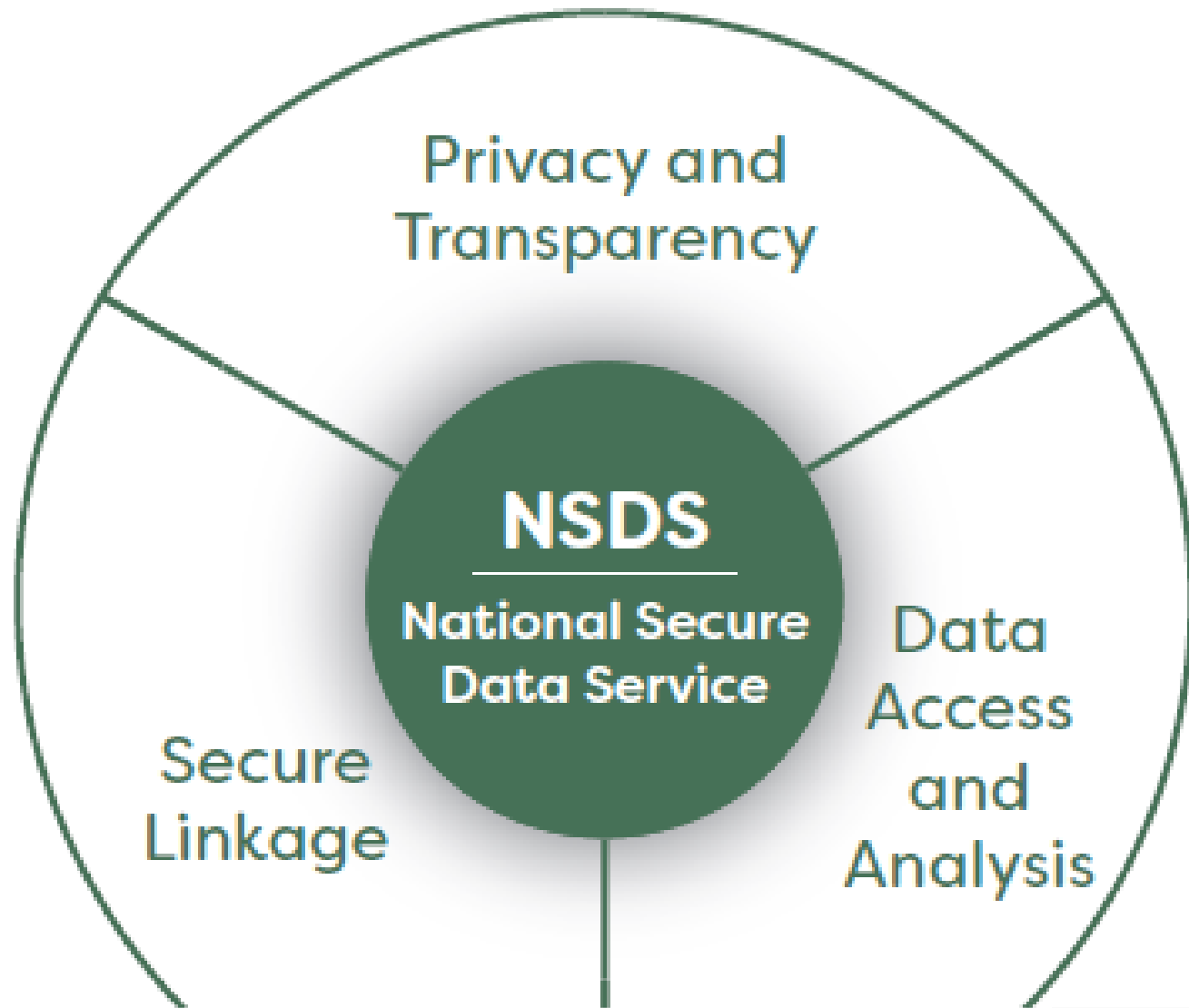


Requires agencies
to make data
accessible in
multiple access
tiers based on data
sensitivity



Requires agencies
to create a “single
front door” for
those who want to
apply for access to
restricted data

Figure 6. Key Functions of the National Secure Data Service (NSDS)



**Functions of
CEP's NSDS**

The Federal Statistical System: What We Do

Barry W Johnson

Director, Statistics of Income

Acting Chief Research and Analytics Officer

Internal Revenue Service

Our Mission: Acquiring, curating, protecting and disseminating data and official statistics to inform the nation

Agencies gather data from many sources in developing official statistics

- Some statistical agencies rely on statistical surveys as the primary source of data for producing official statistics
- Survey data are often collected through partnerships with survey research institutions or other Federal Agencies such as the Census Bureau
- Surveys build on decades of research resulting in sophisticated statistical methods for addressing quality and accuracy challenges
- Advantages include the ability to ensure longitudinal consistency and to collect data using carefully tested, targeted questions that consistency across responses
- Challenges include increasing costs, decreasing response rates and lengthy data collections periods.



Example: Census Bureau's American Community Survey



United States
Census
Bureau

- Premier source for detailed population and housing information about our nation.
- Ongoing survey that provides vital information on a yearly basis.
- Used by public officials, planners, and entrepreneurs use this information to assess the past and plan the future.
- Helps determine how more than \$675 billion in federal and state funds are distributed each year.

Some
programs
combine
survey and
administrative
or other data

- Combining survey and administrative or data from other sources can reduce respondent burden, lower data collection costs and improve timeliness
- Alternate data sources pose different data challenges
- Coverage and content are determined by program needs and may not align with research interests
- Metadata may be suboptimal or difficult to obtain
- Data may need significant additional processing to meet quality and consistency requirements
- Legislative changes in administrative programs can significantly change data content, often with very little notice.
- There is much work going on to develop new statistical methods to address challenges

Programs that combine survey and administrative or other data (cont.)

- Example: National Center for Education Statistics National Postsecondary Student Aid Study
 - Examines the characteristics of students in postsecondary education, with a special focus on how they finance their education
 - Connects multiple data sources including student interviews, institution records, government databases, and other administrative sources
 - State-representative data for most states available from NPSAS:18-AC (expected 2020)
- Example: Census's Longitudinal Employer-Household Dynamics (LEHD)
 - Combines state Unemployment Insurance earnings data and the Quarterly Census of Employment and Wages (QCEW) data with the Bureau's Census and economic survey data.



Some
statistical
programs
solely use
administrative
data: Statistics
of Income

- Data from the more than 250 forms and schedules provided to IRS for tax administration purposes form core of its products
- SOI statistics are primarily based on statistical samples to allow for:
 - Collecting additional information from forms, schedules, and attachments
 - Coding data items to make them statistically useful
 - Reorganizing information to add consistency
 - Performing rigorous quality checks to improve accuracy and reliability
- Release more than 600 products on [IRS.gov/taxstats](https://www.irs.gov/taxstats) annually with more than 10 million downloads



Accessing Tax Data: SOI's Joint Statistical Research Program (JSRP)

- The Tax Code, Title 26, limits who can use tax data and for what purposes
- The JSRP expands opportunities for the research community to work on projects that provide benefit to tax administration
- Researchers gain staff-like access and access data at either an IRS facility or through the Federal Statistical Research Data Centers.
- Outputs result [SOI Working Papers](#), journal articles and conference presentations
 - Chetty, Friedman and Saez on the impact of refundable tax credits, changes in the distribution of income and wealth
 - Estimates of intergenerational income mobility and how retirement assets are used
 - Hoxby on the impact of education tax credits on college attendance
 - Better understanding of gig/contract workers and offshore wealth holdings
 - Behavioral insights tool kit
- Research projects have transformed SOI data systems
- The opportunity to work with respected researchers helps develop and retain talented staff

Through the ICSP and Chief Statistician, We Operate as a System to Advance Individual and Shared Goals

Laws: Paperwork
Reduction Act,
CIPSEA, Evidence Act



- Chief Statistician
- Interagency Council on Statistical Policy
- Federal Committee on Statistical Methodology



Committee on National Statistics



External partnerships
and advisory groups





Federal Statistical Research Data Centers:

A Case Study in Collaborating to Make Data Available to Researchers

- Creating Academic partnerships for more than 30 years
- Increases access to restricted Federal Microdata to:
 - Inform public debate
 - Understand the factors shaping our economy and society
 - Serve the mission of federal and state governments
 - Improve utility and quality of data collections
- Program Management Office housed at Census Bureau manages daily operations and coordination across agencies
 - Staffing
 - Information Technology
 - Physical Security
 - Legal agreements
- Institutional Partners serve as critical link between agencies and researchers
 - Navigate legal requirements
 - Help shape research proposals
 - Provide support through approval process
 - Help coordinate disclosure reviews



Current Brick
and Mortar
FSRDC
Locations

FSRDCs Today

- 6 statistical agencies currently make data available through this system and several others are considering, or are in process of, supplementing their own restricted-access data programs by joining the FSRDCs
- Multi-partner governance structure that includes agencies and institutional partners
- More than 430 projects, although some have paused due to COVID restrictions
 - Virtual Pilot is supporting 38 of these.
 - Pilot is helping to inform physical and cyber security requirements, training and monitoring needs, methods to provide project support from conception to publication
 - Working to improve researcher application process as required under the Evidence Act
- Challenges include long-term funding for technology, staffing and support
- Should be considered a component of a national tiered data access model
- Examples: [Comprehensive Income Dataset](#): a prototype for a restricted micro-level dataset that combines the demographic detail of survey data with the accuracy of administrative measures.

Opportunity Insights: [The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility](#)



Example: Opportunity Insights

[The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility](#)

ARCH

DATA

POLICY

BIG DATA COURSE

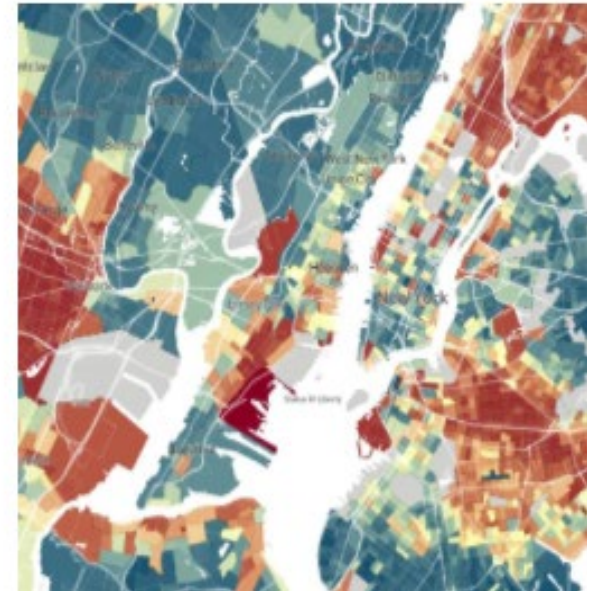
AB



Atlas: Mapping the of Social Mobility

Nathaniel Hendren, Maggie R. Jones, Sonya R.

NO. 25147



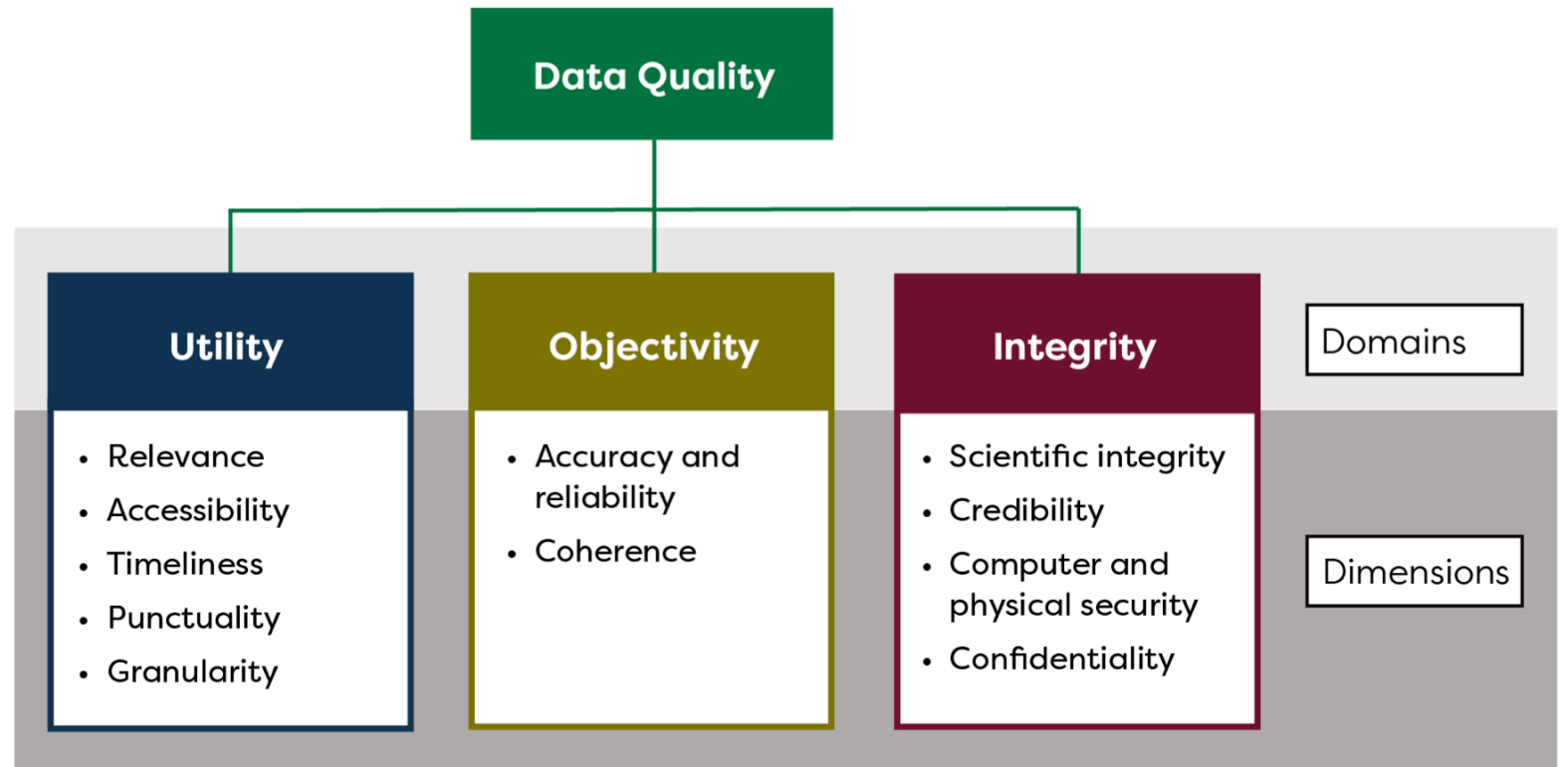
Federal Committee on Statistical Methodology

- Conducts research to help modernize statistical agency practices and policies
 - Made up of senior technical experts from across federal agencies
 - Develops new statistical methods, provides OMB technical recommendations for new or revised standards, advises on best practices of protecting privacy
 - Shares information broadly through annual conferences, workshops, publications and seminars



Example: FCSM – A Framework for Data Quality

Expands traditional view of data quality



Other FCSM Reports on Transparent Reporting of Data Quality products

- Transparent Reporting for Integrated Data Quality: Practices of Seven Federal Statistical Agencies https://nces.ed.gov/FCSM/pdf/Transparent_Reporting_FCSM_19_01_092719.pdf
- Transparent Quality Reporting in the Integration of Multiple Data Sources: A Progress Report, 2017-2018 https://nces.ed.gov/FCSM/pdf/Quality_Integrated_Data.pdf
- Findings from the Integrated Data Workshops hosted by the Federal Committee on Statistical Methodology and Washington Statistical Society https://nces.ed.gov/FCSM/pdf/Workshop_Summary.pdf
- Transparency in the Reporting of Quality for Integrated Data: A Review of International Standards and Guidelines <https://www.mathematica.org/our-publications-and-findings/publications/transparency-in-the-reporting-of-quality-for-integrated-data-a-review-of-international-standards>

Example: FCSM Data Protection Toolkit

Information and Resources for Statistical Officials, Chief Data Officers, and Agency Staff

Disclosure Risk Assessment

- Factors that impact disclosure risk
- Risk assessment “playbook”
- Automated risk assessment tools

Disclosure Risk Mitigation

- SDL methods inventory
- Best practices for various types of data
- Emerging methods and ongoing research

Controlled Access to Sensitive Data

- Inventory of tiered access modes used by federal agencies
- Best practices for controlling risk while promoting access for evidence-building

Assessing Data Accuracy

- Strategies for defining accuracy/quality
- Inventory of accuracy/quality metrics

Governance

- Best practices for establishing a Disclosure Review Board
- Agency policies and charters
- DRB Checklists, Templates, Tools



Find resources to protect your data through Best Practices and Tools as well as a comprehensive search.

LEARN TO USE THE TOOLKIT

SEARCH TOOLKIT RESOURCES

EXPLORE THE TOOLKIT



PROMOTING DATA ACCESS WHILE PROTECTING CONFIDENTIALITY

Federal agencies, both within and outside the statistical system, collect large amounts of data about individuals, businesses, and other organizations.



ASSESSING DISCLOSURE RISK

For an agency to properly protect against disclosure of confidential data, whether through the use of statistical disclosure limitation methods or through tiered data access modes, it must first be able to assess



STATISTICAL DISCLOSURE LIMITATION

The statistical techniques that are used to mitigate the risk of disclosure in data releases are known as Statistical Disclosure Limitation (SDL) methods.

Committee on National Statistics

- Established in 1972 at the National Academies of Sciences, Engineering, and Medicine with cores support from federal statistical agencies
- Provides an independent review of federal statistical activities
- Develops and oversees consensus studies, workshops, and other activities that:
 - Advance the quality of statistical information,
 - Contribute to the statistical policies
 - Provide a forward-looking vision for the federal statistical





In summary, the Federal Statistical System:

- Increasingly relies on data from wide range of sources
- Through a set of federal laws and under the leadership of OMB and the Chief Statistician has a developed an infrastructure to support continual improvement and collaboration
- Collaborates through the ICSP to regularly develop strategic goals that advance federal statistics and support the federal data strategy
- Develops and implements new statistical methods to produce trusted, timely, relevant data for a broad range of users and purposes
- Through the FSRDC network is expanding access to data for research.

The Federal Statistical System: Where We Are Going

Emilda B. Rivers

Director, National Center for Science and Engineering Statistics (NCSES), NSF

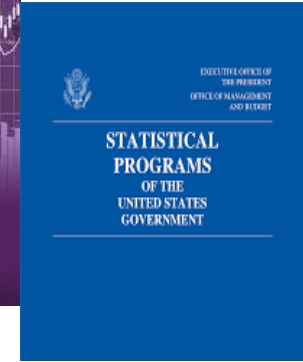
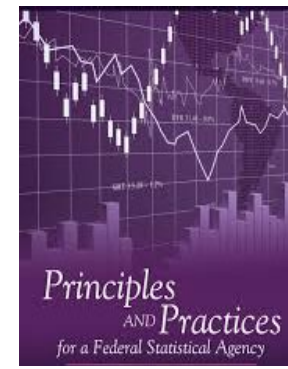
January 22, 2021

Where we are going

- Strategic Priority Areas for the Interagency Council on Statistical Policy (ICSP)
- Data Service Pilot Projects by the Census Bureau and Bureau of Economic Analysis (BEA)
- Data Ecosystem with Evidence Act Responsibilities

ICSP Strategic Priority Areas

- Data quality best practices
- Autonomy and objectivity issues
- **Enabling use of blended and high frequency data for statistical purposes**
- ✓ Evidence Act regulations and processes (focus on Standard Application Process)
- ✓ Establishing agency Statistical Official role
- ✓ Collaborating with Chief Data Officers
- Developing staff data skills
- Data confidentiality/re-identification risk
- ✓ Establishing data service vision



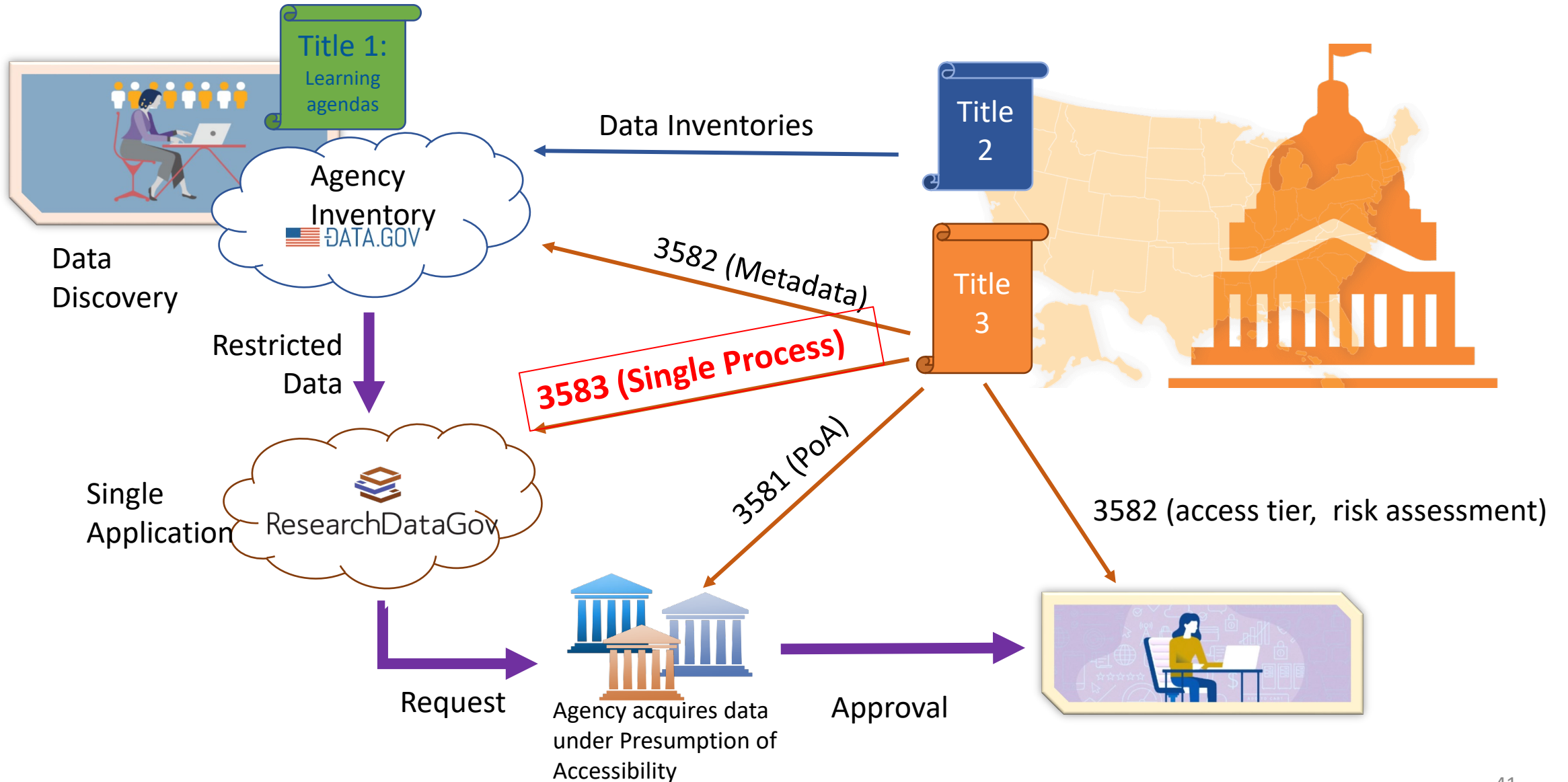
✓ **Explicit connection to the Evidence Act**

Data Service Priority Areas By Census And BEA

Priority Areas

- ✓ Privacy research
- ✓ Supporting improved access and use of data by external stakeholders
- ✓ Acquiring non survey-based data sources and blending existing survey and administrative data
- ✓ Creating beta products to evaluate key data users and stakeholders

Data Ecosystem with Evidence Act Responsibilities



The Law - 3583. Application To Access Data Assets For Developing Evidence

(a) **STANDARD APPLICATION PROCESS.** The Director shall establish a process through which agencies, the Congressional Budget Office, State, local, and Tribal governments, researchers, and other individuals, as appropriate, may apply to access the data assets accessed or acquired under this subchapter by a statistical agency or unit for purposes of developing evidence. The process shall include the following:

- (1) Sufficient detail to ensure that each statistical agency or unit establishes **an identical process**.
- (2) **A common application form**.
- (3) Criteria for statistical agencies and units to **determine whether to grant an applicant access** to a data asset.
- (4) Timeframes for prompt determinations by each statistical agency or unit.
- (5) **An appeals process** for adverse decisions and noncompliance with the process established under this subsection.
- (6) **Standards for transparency**, including requirements to make the following information publicly available:
 - (A) Each application received.
 - (B) The status of each application.
 - (C) The determination made for each application.
 - (D) Any other information, as appropriate, to ensure full transparency of the process established under this subsection.

(b) **CONSULTATION.**—In establishing the process required under subsection (a), the Director shall consult with stakeholders, including the public, agencies, State and local governments, and representatives of non-governmental researchers.

(c) **IMPLEMENTATION.**—The head of each statistical agency or unit shall implement the process established under subsection (a).”

STANDARD APPLICATION PROCESS (SAP)



The image shows the homepage of the ResearchDataGov website. At the top, there is a dark blue navigation bar with white text links: "LOG IN/CREATE ACCOUNT", "BROWSE BY AGENCY -", "BROWSE ALL DATA", "ABOUT", "CONTACT", and "FAQ". Below the navigation bar is a large, light beige banner with a subtle pattern of concentric white arcs. In the center of the banner is the ResearchDataGov logo, which consists of three stacked, stylized rectangular blocks in orange, blue, and orange. Below the logo, the text "ResearchDataGov" is written in a large, black, sans-serif font. Underneath the main title, the subtitle "Application Portal for Restricted Data for Federal Statistics" is displayed in a smaller, dark blue font. At the bottom of the banner, there is a white search bar with a thin orange border. Inside the search bar, the placeholder text "enter your query" is written in a grey font, and a magnifying glass icon is positioned to the right of the text.

SAP Development Phases

Phase 1

- Ended 08/2020
- Pilot by Census with Michigan

Phase 2

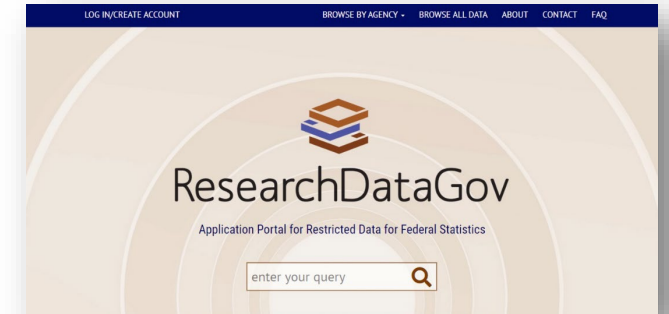
- 08/2020-08/2021
- ICSP technical and policy development and requirements gathering

Phase 3

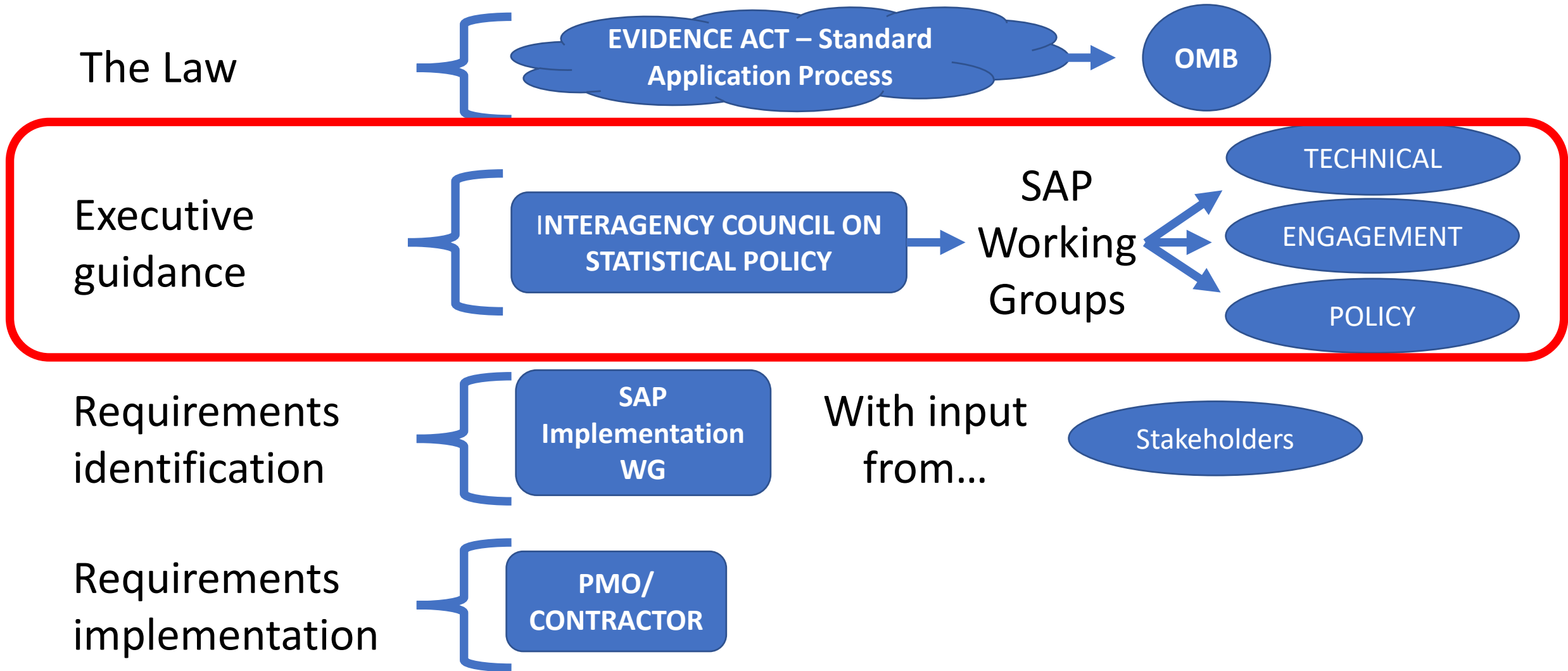
- 08/2021-02/2022
- Implementation of enhanced portal

Phase 4

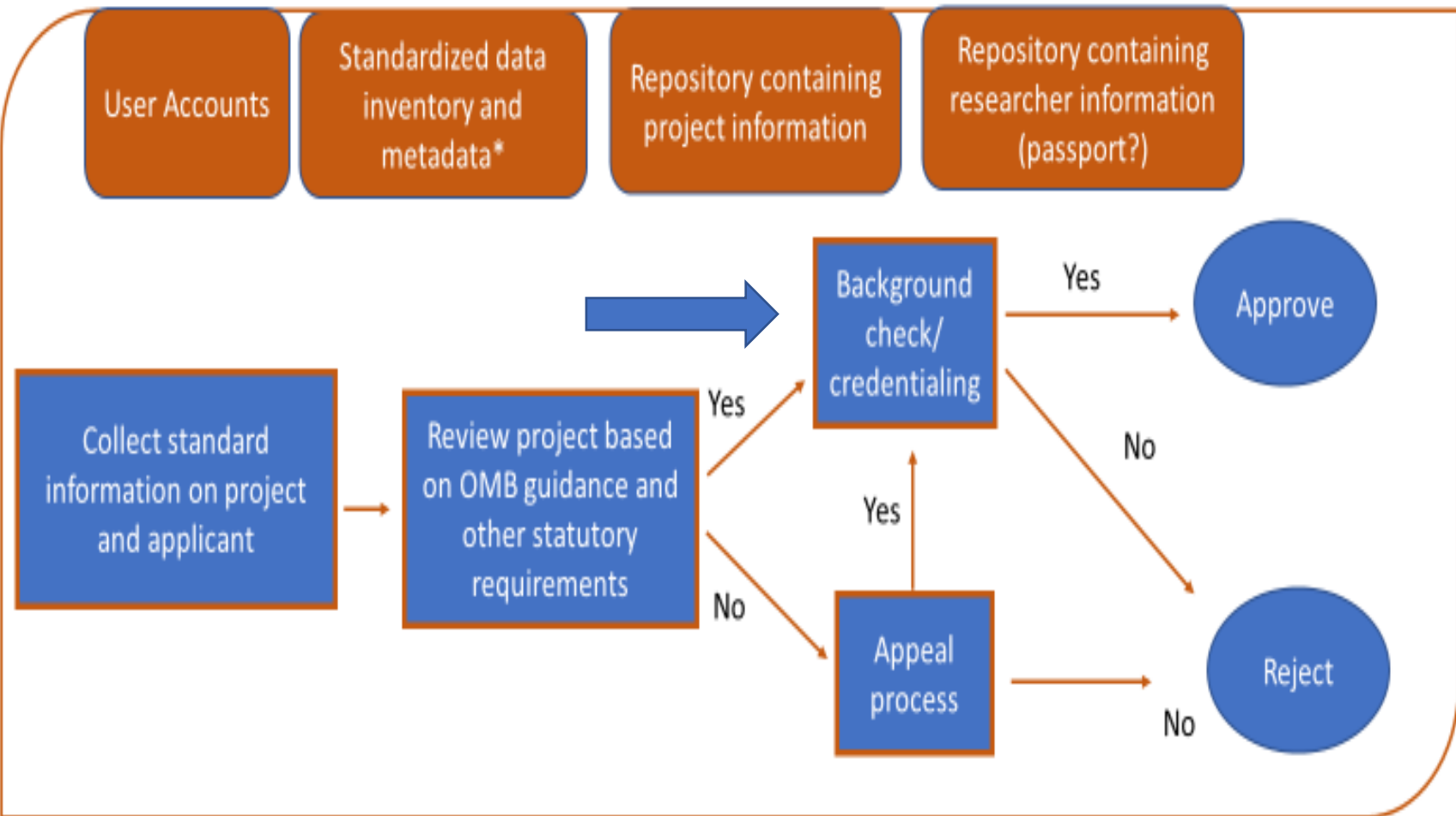
- 02/2022
- Application in full development mode
- Maintenance of the application



SAP Development Process



SAP Framework



Tracking Process so applicants can monitor approval progress

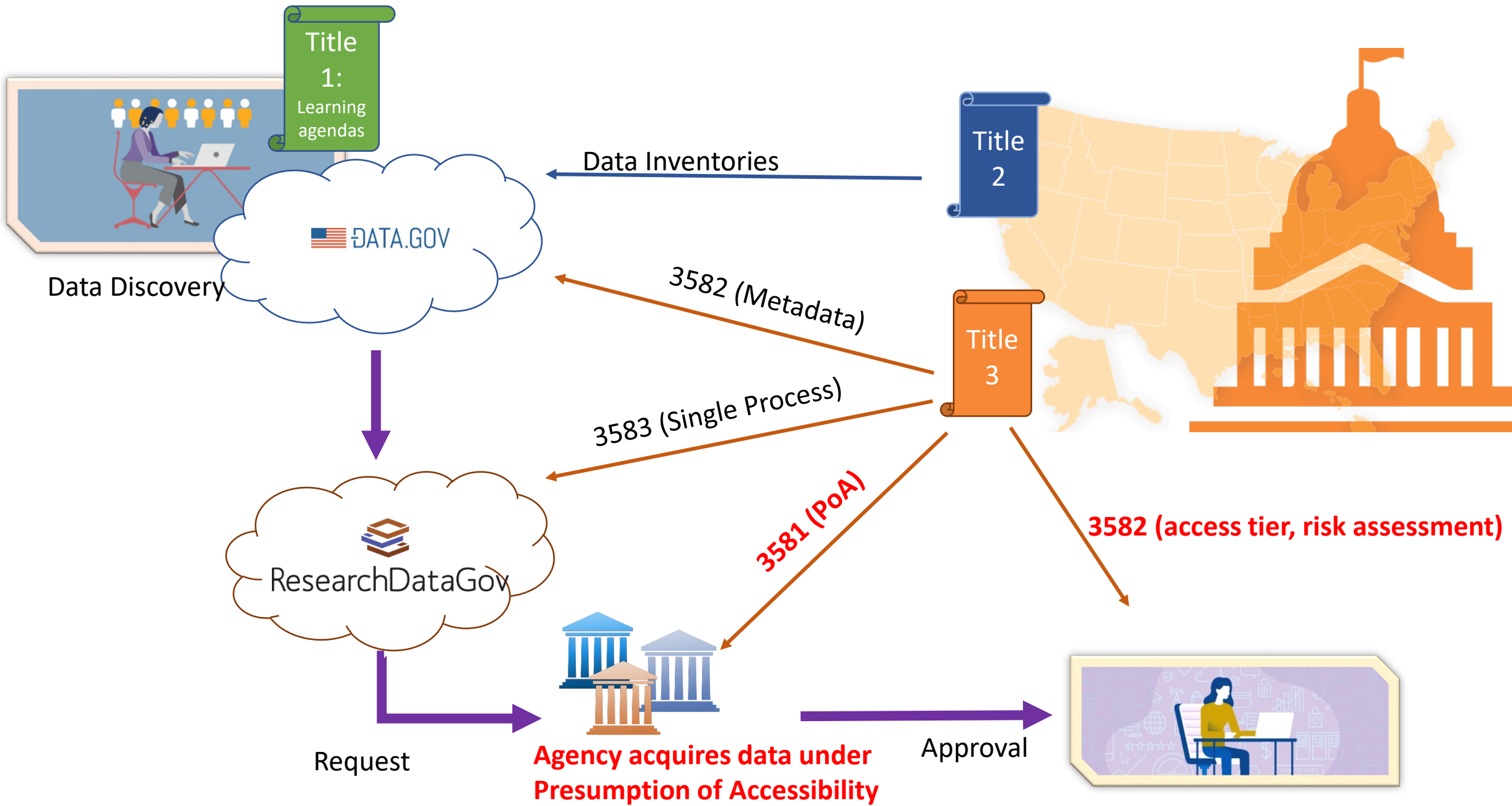
*Ideally this information would reside on agency websites and be accessed by the standard application

Components (in blue)

- Web-based application
- Project review
- Appeals process
- Background check

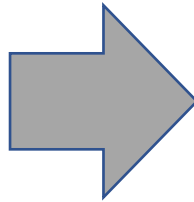
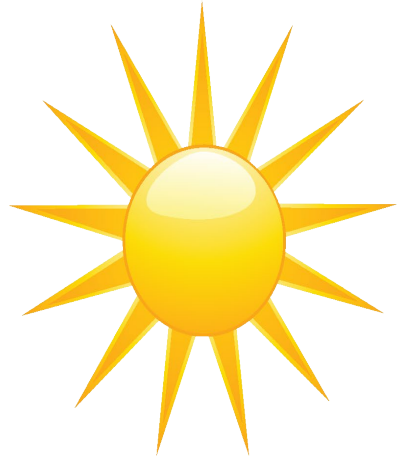
Support Database (in orange)

- User accounts
- Project information
- Researcher information
- Metadata

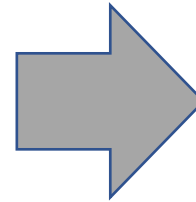
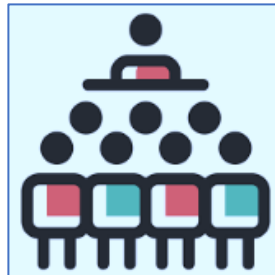
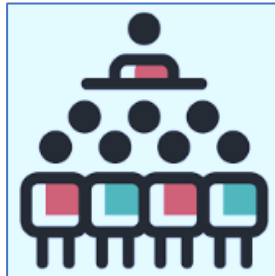
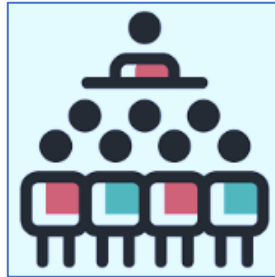


Data Ecosystem with Evidence Act Responsibilities

**Statistical Agency
with Legal Authority**



Data Sources



SECURE DATA SERVICE

*Improving the ability to acquire, access
and combine data through ...*



**Modernizing data
infrastructure**



**Spurring
innovation**



**Building
partnerships**

Time for Break then Questions !

