The U.S. Approach to Leveraging National Data As a Strategic Asset

Dr. Nancy Potok
Data Symposium
Wellington,
March 9, 2023

Overview

U.S. Data Ecosystem
Federal Data Strategy
Evidence Based Policy
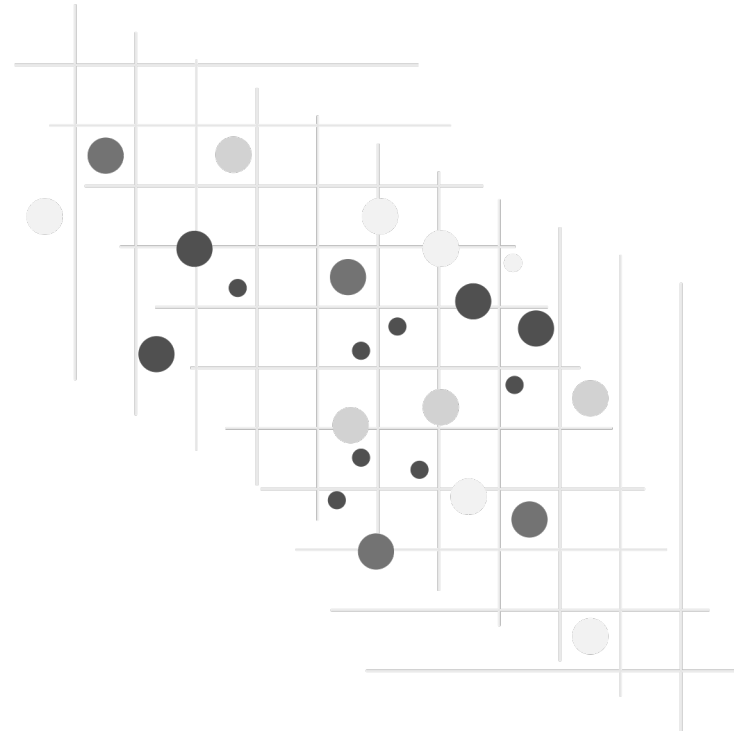Administrative Policies and Guidance to Agencies
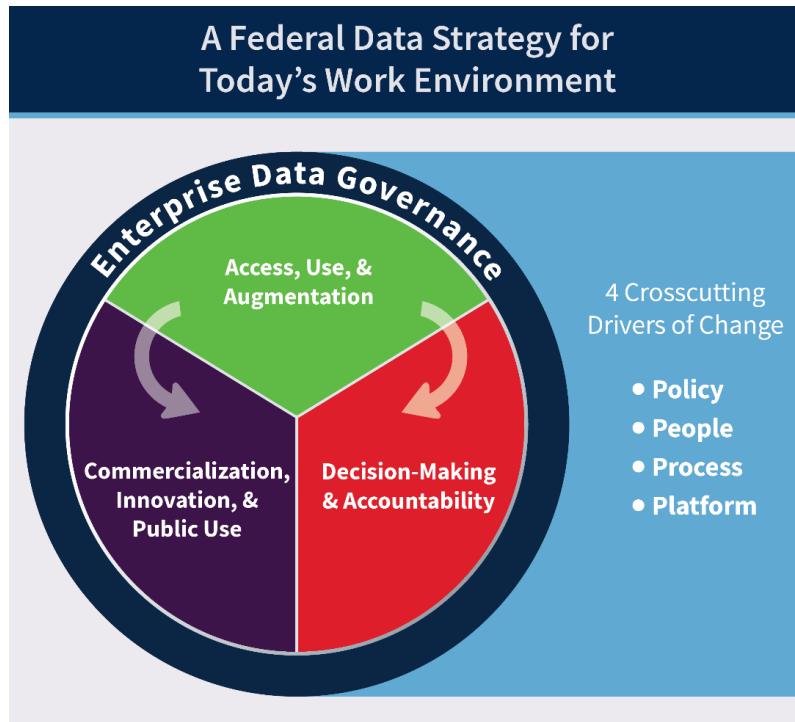Infrastructure
Tools

Data Ecosystem

Federal Data Strategy

Foundations of Evidence Based Policy Act

CHIPS and Science Act

Administrative Actions

Data Providers

Data Guardians

Infrastructure Owners

Data Users

Policy Makers

Federal Data Strategy: Leveraging Data as a Strategic Asset

A Federal Data Strategy for Today's Work Environment

Enterprise Data Governance

- Access, Use, & Augmentation
- Commercialization, Innovation, & Public Use
- Decision-Making & Accountability

4 Crosscutting Drivers of Change

- Policy
- People
- Process
- Platform

Achievement of the goal will leverage four high-level areas, supported by four cross-cutting drivers of change.
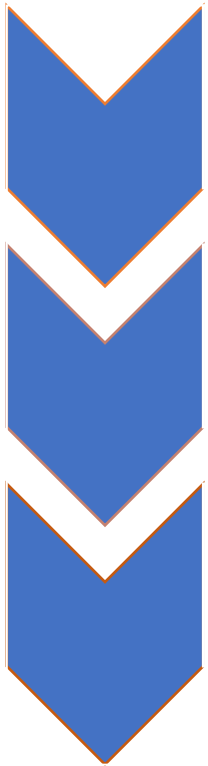
Federal Data Strategy Components

GUIDING PRINCIPLES

BEST PRACTICES

ANNUAL ACTION PLAN

10 Guiding Principles: ETHICAL GOVERNANCE

1. **Uphold Ethics**: Monitor and assess the implications of federal data practices for the public. Design checks and balances to protect and serve the public good.

2. **Exercise Responsibility:** Practice effective data stewardship and governance. Employ sound data security practices, protect individual privacy, maintain promised confidentiality, and ensure appropriate access and use.

3. **Promote Transparency:** Articulate the purposes and uses of federal data to engender public trust. Comprehensively document processes and products to inform data providers and users.

10 Guiding Principles: CONSCIOUS DESIGN

4. Ensure Relevance: Protect the quality and integrity of the data. Validate that data are appropriate, accurate, objective, accessible, useful, understandable, and timely if possible and acquire additional data if needed

5. Harness Existing Data: Identify data needs to inform priority research and policy questions; reuse data

6. Anticipate Future Uses: Create data thoughtfully, considering fitness for use by others; plan for reuse and build in interoperability from the start

7. Demonstrate Responsiveness: Improve data collection, analysis, and dissemination with ongoing input from users and stakeholders. The feedback process is cyclical; establish a baseline, gain support, collaborate, and refine continuously.

10 Guiding Principles: LEARNING CULTURE

8. **Invest in Learning**: Promote a culture of continuous and collaborative learning with and about data through ongoing investment in data infrastructure and human resources

9. **Develop Data Leaders**: Cultivate data leadership at all levels of the federal workforce by investing in training and development about the value of data for mission, service, and the public good

10. **Practice Accountability**: Assign responsibility, audit data practices, document and learn from results, and make needed changes

Statutes - Evidence Act of 2018

Foundations for Evidence-Based Policy Making Act of 2018

Recommendations from the U.S. Commission on Evidence Based Policy Making

Three Titles:

    I.   Federal Evidence Building Activities

    II.   Open Government Data Act

    III. Confidential Information Protection and Statistical Efficiency Act

.

PUBLIC LAW 115–435—JAN. 14, 2019   132 STAT. 5529
Public Law 115–435 115th Congress

An Act

To amend titles 5 and 44, United States Code, to require Federal evaluation activi- ties, improve Federal data management, and for other purposes.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.
SHORT TITLE.—This Act may be cited as the "Foundations for Evidence-Based Policymaking Act of 2018".
TABLE OF CONTENTS.—The table of contents for this Act is as follows:
Sec. 1. Short title; table of contents.

TITLE I—FEDERAL EVIDENCE–BUILDING ACTIVITIES
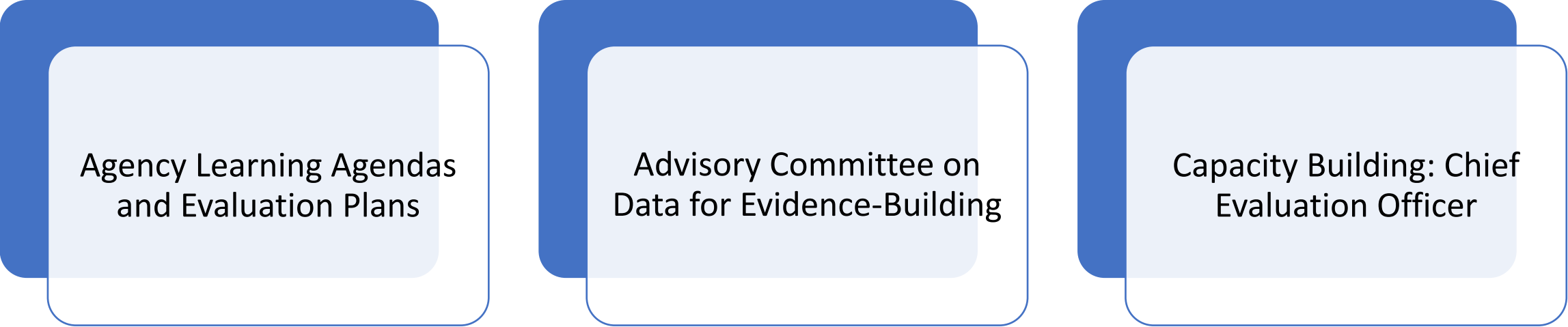Sec. 101. Federal evidence-building activities.

TITLE II—OPEN GOVERNMENT DATA ACT
Sec. 201. Short title.
Sec. 202. OPEN Government data.

TITLE III—CONFIDENTIAL INFORMATION PROTECTION AND STATISTICAL EFFICIENCY

Sec. 301. Short title.
Sec. 302. Confidential information protection and statistical efficiency. Sec. 303. Increasing access to data for evidence.

TITLE IV—GENERAL PROVISIONS
Sec. 401. Rule of construction.
Sec. 402. Use of existing resources. Sec. 403. Effective date.

5 USC 301 prec.
Foundations for Evidence-Based Policymaking Act of 2018.
5 USC 101 note.
Jan. 14, 2019
    [H.R. 4174]

Title 1 – Federal Evidence-Building Activities

Agency Learning Agendas and Evaluation Plans

Advisory Committee on Data for Evidence-Building

Capacity Building: Chief Evaluation Officer

M-21-27 Learning Agendas

"OMB expects agencies to use evidence whenever possible to further both mission and operations, and to commit to build evidence where it is lacking...

OMB strongly believes that implementing the Evidence Act is not a compliance exercise, and that agencies should develop the required Title I deliverables (i.e., the Learning Agenda, Annual Evaluation Plan, and Capacity Assessment for Statistics, Evaluation, Research and Analysis) in a way that fulfills their purpose as strategic, evidence-building plans...".

https://www.evaluation.gov/evidence-plans/learning-agenda/

Health and Human Services Learning Agenda

The activities in this priority area include increasing affordability and enrollment in high quality healthcare coverage; equitable access to quality health-care services; integrating behavioral healthcare into the healthcare system; expanding access to mental health and substance use disorder treatment and recovery services; and bolstering the healthcare workforce.

Healthcare Priority Questions

- How do HHS policies and programs increase choice, affordability and enrollment in high-quality healthcare coverage?
- To what extent do HHS programs and policies reduce costs and improve quality of healthcare services?
- How and to what extent do HHS programs and policies ensure access to safe medical devices and drugs?
- How do HHS programs and policies expand equitable access to comprehensive, community-based, innovative, and culturally-competent healthcare services while addressing social determinants of health?
- How effective are HHS programs and policies at integrating behavioral health services into the healthcare system?
- To what extent do HHS programs and policies strengthen and expand access to mental health and substance use disorder treatment and recovery services for individuals and families?
- How do HHS programs and policies bolster the primary and preventive healthcare workforce to ensure delivery of quality services and care?

Healthcare Evidence-Building Activities

Divisions across HHS are conducting evidence-building activities to address these questions, including ASPR, ACL, CDC, CMS, NIH, and SAMHSA. These activities address various HHS programs, such as Medicare, Medicaid, Community Health Centers, and grantmaking programs like the National Paralysis Resource Center (NPRC). Eight identified evidence-building activities address healthcare priority questions. Full details on these activities are provided in the appendix.

Populations impacted by the healthcare evidence-building activities contained in this plan include mothers, individuals with disabilities (physical and cognitive disabilities), children, homeless individuals, and those recovering from substance use disorders. Notably, there are several instances of multiple divisions building evidence around the same populations and healthcare topics. An example includes the NIH impact assessment for The Role of Opioids in the Treatment of Chronic Pain Pathways to Prevention Workshop and the CMS evaluation of Maternal Opioid Misuse (MOM) Model. Additionally, some evidence-building activities that address the healthcare priority questions may seek to improve services, assess resource center effectiveness, advance telehealth capabilities, bolster preventative care, and identify promising value-based insurance models.

HHS executes a broad range of evidence-building activities to answer these priority questions. Activities include but are not limited to policy analysis, descriptive analysis, program evaluation, foundational fact-finding, and performance measurement. Most of the activities in this plan use a combination of methods to address a priority question. For example, using administrative data and program participant interviews to evaluate telehealth strategies to address hypertension management and control or site visits along with claims analysis to evaluate the Maternal Opioid Misuse Model. The activities utilize

Page 4 of 44

12

Statutes – One Application Mandate

The Standard Application Process

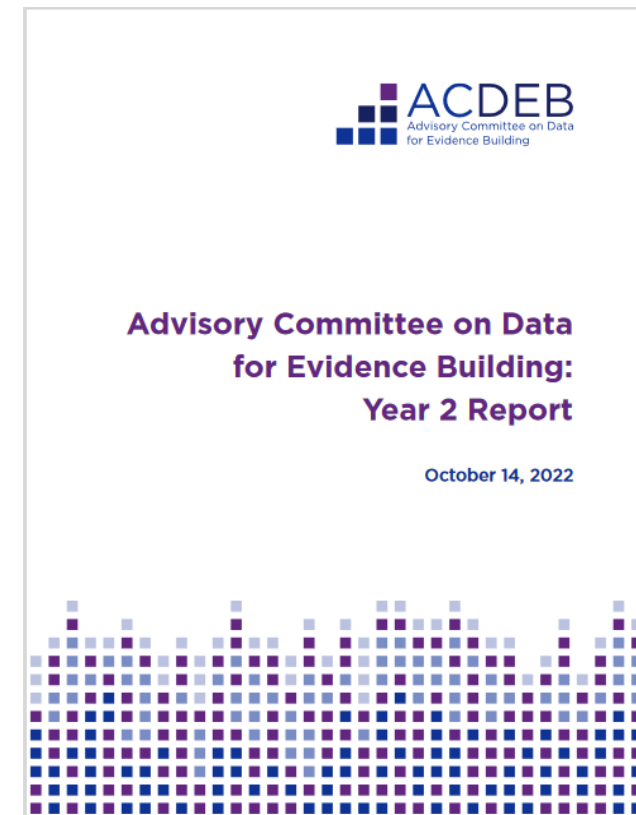|	
The Standard Application Process	
Background and Overview	
Roles and Responsibilities	
Benefits of the SAP	
Confidentiality and Privacy	
How to Get Involved	
Phases of Development	+
Glossary	
Frequently Asked Questions	

The federal statistical system has adopted a standard application process (SAP) for applying for access to confidential data assets from the nation's statistical agencies. The SAP marks an important milestone for the federal statistical system. For the first time, primary statistical agencies and units have coordinated and agreed to use the same application for access to their restricted-use data assets.

Statute – National Advisory Committee

Evidence Act: Implementation Recommendations for a

National Secure Data Service

ACDEB
Advisory Committee on Data
for Evidence Building

**Advisory Committee on Data
for Evidence Building:
Year 2 Report**

October 14, 2022

Infrastructure: Evidence Commission Recommendation

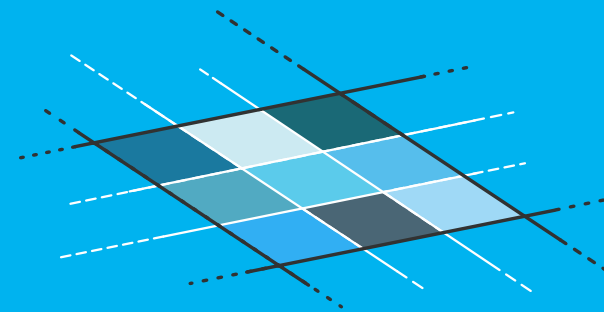Data linkage, not data warehouse

Enhanced privacy protections

Center of expertise

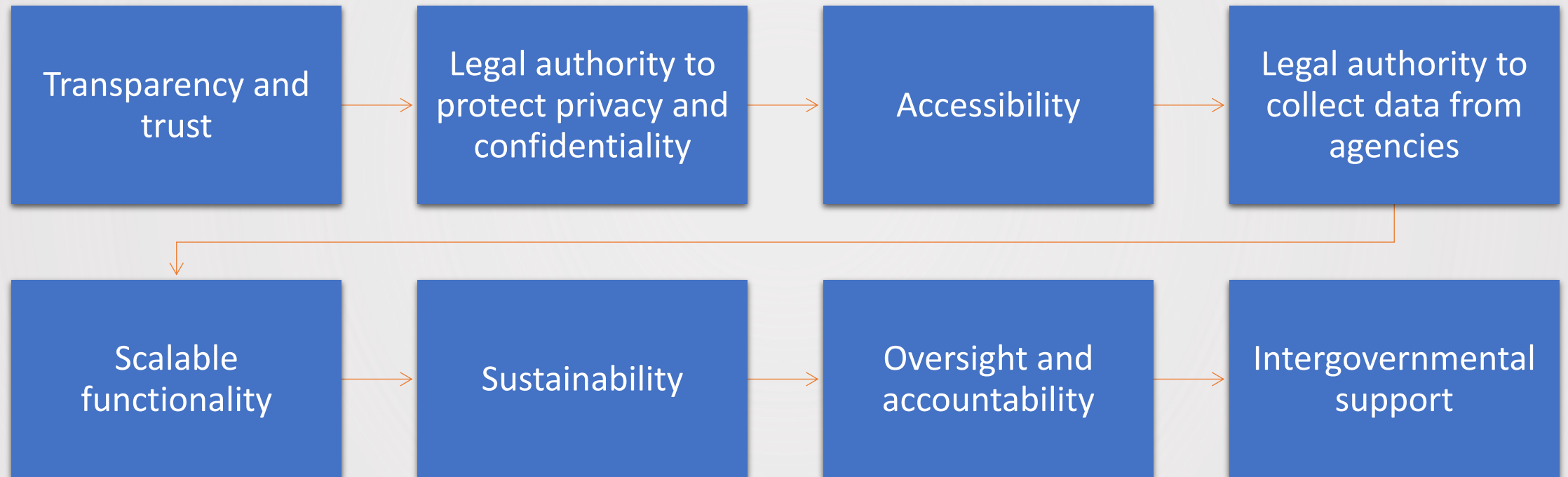# Designing a National Secure Data Service

White Paper to Inform the Advisory Committee

MODERNIZING U.S. DATA INFRASTRUCTURE:

Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building

DATA FOUNDATION

Datafoundation.org

Necessary Attributes of a Data Service

Transparency and trust → Legal authority to protect privacy and confidentiality → Accessibility → Legal authority to collect data from agencies

Scalable functionality → Sustainability → Oversight and accountability → Intergovernmental support

Scalable Functionality

Increase capacity – Not create bottlenecks

Examples of what is scalable

- Standards setting for interoperability and quality measurement for linked data
- Methods toolbox for preventing reidentification
- Methods toolbox for protecting privacy
- Streamlined application approval process for federal data
- Searchable data inventories
- Accessible researcher portal for info on other research, data quality, etc.

Infrastructure Recommendation

Establish a Federally-Funded Research and Development Center (FFRDC) through the National Science Foundation

A Blueprint for Implementing the National Secure Data Service:

Initial Governance and Administrative Priorities for the National Science Foundation

June 2022

DATA FOUNDATION

Follow up paper to *Modernizing U.S. Data Infrastructure*, published by The Data Foundation (Nick Hart co-author)

Provides a specific blueprint for NSDS implementation, including structure, governance, and activities in several key areas

Recognizes the core activities of an NSDS include hosting a secure infrastructure where researchers would: (1) submit proposed research projects for approval; (2) link and access data for research and analyses; and (3) have research results privacy protected then prepared for public dissemination

Offers options for how the NSDS might get started around its core functions to enable later expansion, but most importantly offers a blueprint for how NSF can begin this work quickly, efficiently, and transparently.

Vision for the NSDS: Coordinating Data Hub

Research Node

Hub that is part of a network in which it plays the central role of coordinating a group of nodes that make up a collaborative system guided by a coherent body and governance process.

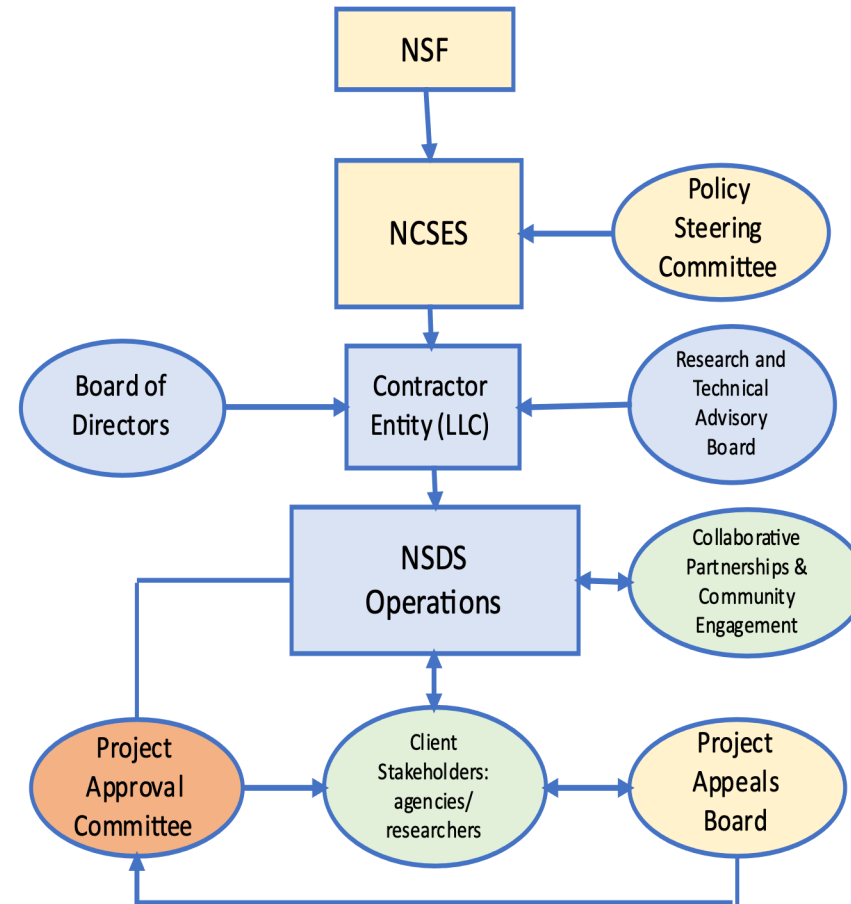Research Node

Research Node

Research Node
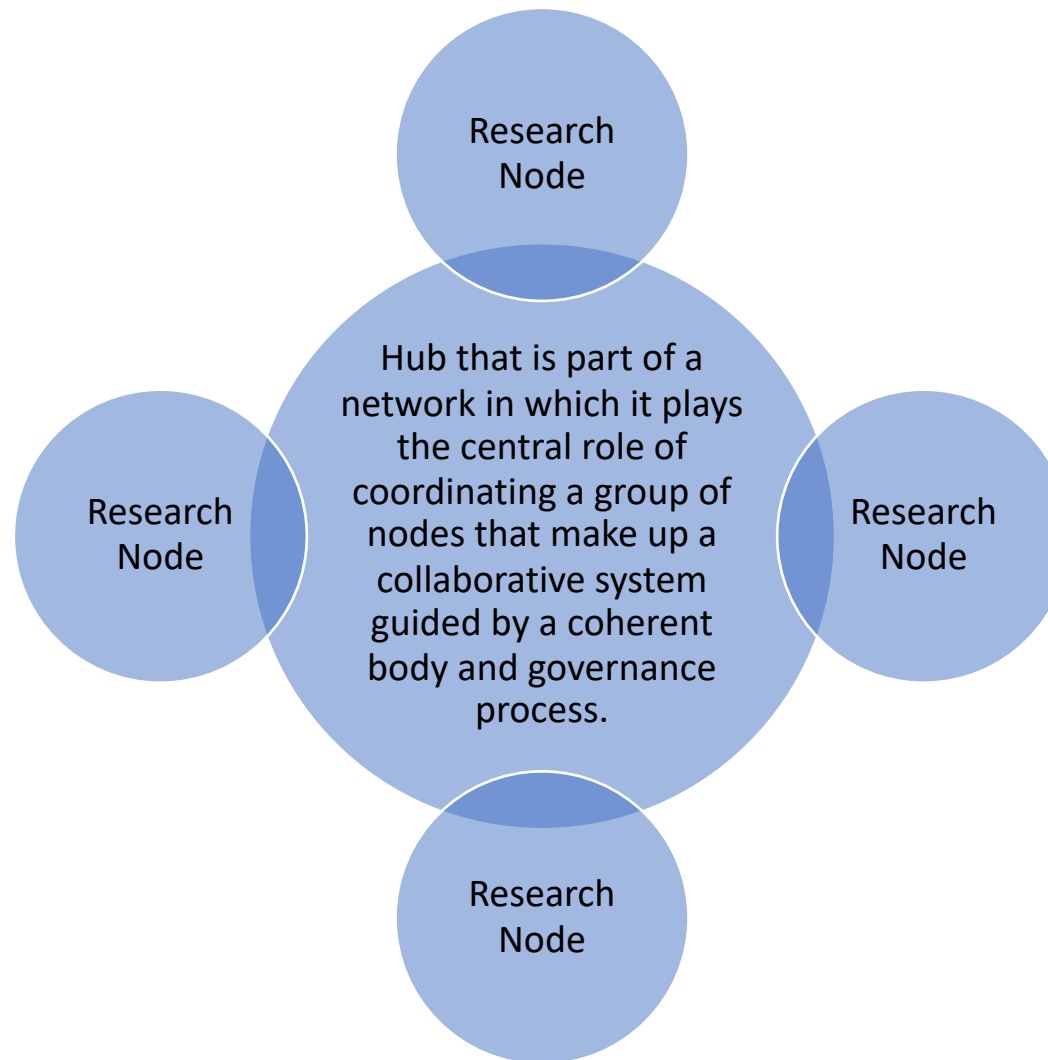
Important Infrastructure Attributes

Efficient, transparent, and timely business processes are essential elements for a successful NSDS. The NSDS should not become a bottleneck

- host the application and approval process; enforce standards

- provide core services

- continue to develop new tools for privacy protection, search and discovery of data usage, quality measurement

- Serve as a catalyst for collaboration

- identify where agency policies are creating bottlenecks and assist agencies in finding more efficient approaches to meeting their requirements

"**Governance** and **oversight** are key to establishing transparency and trust. The NSDS needs to demonstrate to key stakeholders and oversight bodies that it is operating according to broadly accepted principles and practices that are ethical and equitable, and that the work being produced is independent and of high quality. "

Government Owned Contractor Operated
Recommended Governance Model

NSF

NCSES

Policy Steering Committee

Board of Directors

Contractor Entity (LLC)

Research and Technical Advisory Board

NSDS Operations

Collaborative Partnerships & Community Engagement

Project Approval Committee

Client Stakeholders: agencies/ researchers

Project Appeals Board

National Advisory Committee Report

ACDEB
Advisory Committee on Data for Evidence Building

Advisory Committee on Data for Evidence Building: Year 2 Report

October 14, 2022

Recommendations Part 3. NSDS Functions

Function 1. Coordination—Evidence Act Implementation for the Federal Government and Beyond

3.1	The NSDS should coordinate with the ICSP, communities of practice, and other key stakeholder groups to establish best practices for implementing the requirements of the Evidence Act and should serve as a model for testing and demonstrating those best practices for governments at all levels.

Function 1. Coordination—Support a High-Quality User Experience, Including Providing Technical Assistance

3.2	The NSDS website should serve as a "front door" to the nation's data assets, organized around a set of personas that reflect basic user needs.
3.3	The NSDS should identify opportunities for automation of its "intake process," providing a high-quality user experience while focusing staff effort on complex user needs.
3.4	The NSDS should employ data concierges to help users refine their research projects, discover relevant data, and acquire access to that data.
3.5	The NSDS should employ technical assistance leads who develop educational resources for data providers and data users related to the methods and technologies used by the NSDS and in the broader evidence-building ecosystem.
3.6	The NSDS should actively encourage the development of communities to crowdsource support for users, complementing the work of the data concierges.
3.7	The NSDS should regularly sponsor projects that demonstrate the value of streamlining data sharing and increasing coordination, specifically projects that highlight cross-functional, cross-agency, and cross-governmental topics.

Function 2. Communication

3.8	OMB should adopt a clear statement of purpose for the NSDS that is rooted in its core value.
3.9	Through the NSDS website, the National Center for Science and Engineering Statistics (NCSES) should create a public information hub that serves as a central repository for

CHIPS and Science Act of 2022: SEC. 10375 National Secure Data Service

Authorizes NSF to establish a National Secure Data Service pilot

Charged with developing, refining, and testing models to inform a full service

Authorizes USD $9 million per year 2022-2026 (USD $45 million, NZD $75 million)

Requires pilot be set up within one year of enactment

America's Data Hub could be home for the pilot

Role of Academia and Think Tanks

Advance public policy-related research through a connected, collaborative network of ideas and infrastructure:

- Encourage local government agency and NGO collaborations to link and analyze data
- Create more inclusive pipeline of future researchers both inside and outside government
- Create feedback loops between agencies and researchers on quality, accessibility, equity

Research Topics:

- Policy Evidence Building/Evaluation
- Privacy protection
- Advances in AI, Machine Learning, Natural Language Processing, and computer science
- Quality measures development

Local and Regional Government Interests

Improve outcomes for the public:

Increase data sharing across government boundaries

Improve federal reporting requirements

Conduct program evaluations

Improve program operations

Tools: Search and Discovery Platform

Common Goals

Combined Approach

1. Basic Usage Information

How much are agency datasets used in research and how has that usage changed over time?

How often is each one of an agency's identified dataset used in research and how has that usage changed over time?

2. The Agency's Portfolio

What topics are an agency's datasets being used to study and what publications are associated with each topic?

What topics is each one of an agency's identified dataset used to study in research and what publications are associated with each topic?

What other datasets are being used to study each topic?

3. Drilling Into the Details for Each Dataset

Who are the main authors using each agency's datasets? Who are the main authors using each specific dataset?

What are the publications associated with each author?

What institutions are the centers of use for each agency dataset and in what geographic locations are the institutions located?

Tools: Combined Approach

Three Access Points

Developed Together

Democratizing Data - USDA by Democratizing Data

USDA

DATASETS	PUBLICATIONS	AUTHORS	COUNTRIES	CITATIONS	INSTITUTIONS
3	1,752	7,572	58	14,626	4,464

Select a Dataset to Explore Usage | Datasets:All, Year:2017, 2018, 2019 and 3 more | CLEAR FILTERS

Name	Pub	Cit
RUCC	1,033	10,335
NASS Census of Agriculture	666	3,798
Agricultural Resource Managem..	89	632

Filter by Year(s)

Year	Pub	Cit	Authors
2017	169	2,801	667
2018	256	4,146	1,037
2019	320	3,693	1,427
2020	359	2,862	1,557
2021	505	1,100	3,515
2022	143	24	678

Filter by Topic(s)

1,752 Publications — DOWNLOAD SPREADSHEET

Publication	Cit
CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2011-2015	968
CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2..	793
Acceptability of a COVID-19 vaccine among adults in the United States: How many people would get vaccinated?	712
Survival after minimally invasive radical hysterectomy for early-stage cervical cancer	325
CBTRUS statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2013-2017	264
Incidence and prognosis of patients with brain metastases at diagnosis of systemic malignancy: A population-based study	223
Rural-Urban differences in cancer incidence and trends in the United States	143
Brain metastases in newly diagnosed breast cancer: A population-based study	134

910 Journals — DOWNLOAD SPREADSHEET

Publication Title	Pub	Cit
Journal of Rural Health	52	405
Journal of Soil and Water Conservation	38	114
International Journal of Environmental Research and Public Health	32	145
Sustainability (Switzerland)	30	171
Applied Economic Perspectives and Policy	24	191

4,464 Institutions — DOWNLOAD SPREADSHEET

Institution Name	Pub	Cit
RAND Corporation	19	188
Department of Agricultural and Resource Economics, Colorado State University	14	41
College of Nursing, University of Kentucky	11	35
Department of Agricultural and Resource Economics, University of Tennessee	11	18
University of North Carolina at Chapel Hill	11	86
Department of Agricultural Economics, Purdue University	10	78
Department of Agricultural Economics, Kansas State University	9	30
Department of Sociology, Iowa State University	8	84
Holden Comprehensive Cancer Center, University of Iowa	8	9
Department of Agricultural Economics and Economics, Montana State University	7	7

File   Edit   View   Insert   Cell   Kernel   Widgets   Help

Trusted | Py

Retrieve all agency r

The queries below search for data
Here we return that whole table to

```
In [67]: sql="select * from agency_run
         agency_run=cj.executeQuery(s
         agency_run
```

Out[67]:

id

What topics are an agency's datasets being used to study?

```
In [64]: sql=f"""
         with a as (
         select ds.id as ds_id, max(ds.alias) as dataset
         ,       t.id as topic_id, max(t.keywords) as topic
         ,       count(distinct p.id) as num_topic
         ,       rank() over(partition by ds.id order by count(distinct p.id) desc) as rank
            from agency_run ar
```

1) Basic Usage Information

How much are agency datasets used in research and how has that usage changed over time? How often is each one of an agency's identified dataset used in research and how has that usage changed over time?

In [60

How much are agency datasets used in research? or How often is each one of an agency's identified dataset used in research?

The table shows the use of each dataset based on publications that have used respective dataset.

```
In [61]: sql=f"""                                        -- using a python f-string so that parameters AGENCY and VERSION can
         select ds.alias as dataset, p.year
         ,       count(distinct d.publication_id) as pub_per_year    -- count the different publications
           from agency_run ar                            -- the table storing metadata about the individual runs executed by
           join dyad d on d.run_id=ar.id                  -- JOIN to the table with all the dyads
           join publication p on p.id=d.publication_id    -- JOIN to the table with publications
           join dataset_alias da on d.dataset_alias_id = da.id       -- first JOIN to the dataset_alias table with the *aliases* of the t
           join dataset_alias ds on ds.alias_id = da.parent_alias_id  -- second JOIN to retrieve the actual dataset, defined as the *paren
         where ar.agency='{AGENCY}' and ar.version='{VERSION}'      -- restriction of the agency run to the AGENCY/VERSION defined above
         group by ds.id, ds.alias, p.year                 -- we group on the dataset and the year to count distinct publicatio
         order by dataset,year
```

Future Steps for Exploration

- Generate topic-based analysis of coverage in priority areas and inform future investments

- Create a cross agency view of data sharing and linkage activities across the federal statistical system

- Use the platform to facilitate routine engagement with the research, program and policy communities and to build collaborators

- Utilize synergies within existing federal requirements such the Standard Application Process and the National Secure Data Service demonstration project to build on open access and transparency efforts
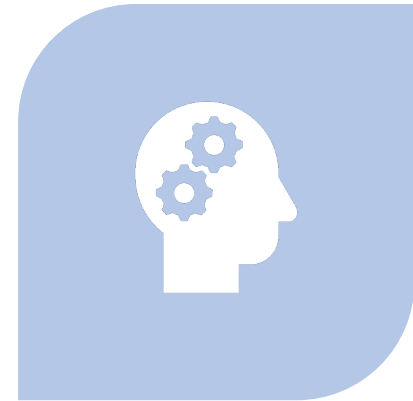
Website
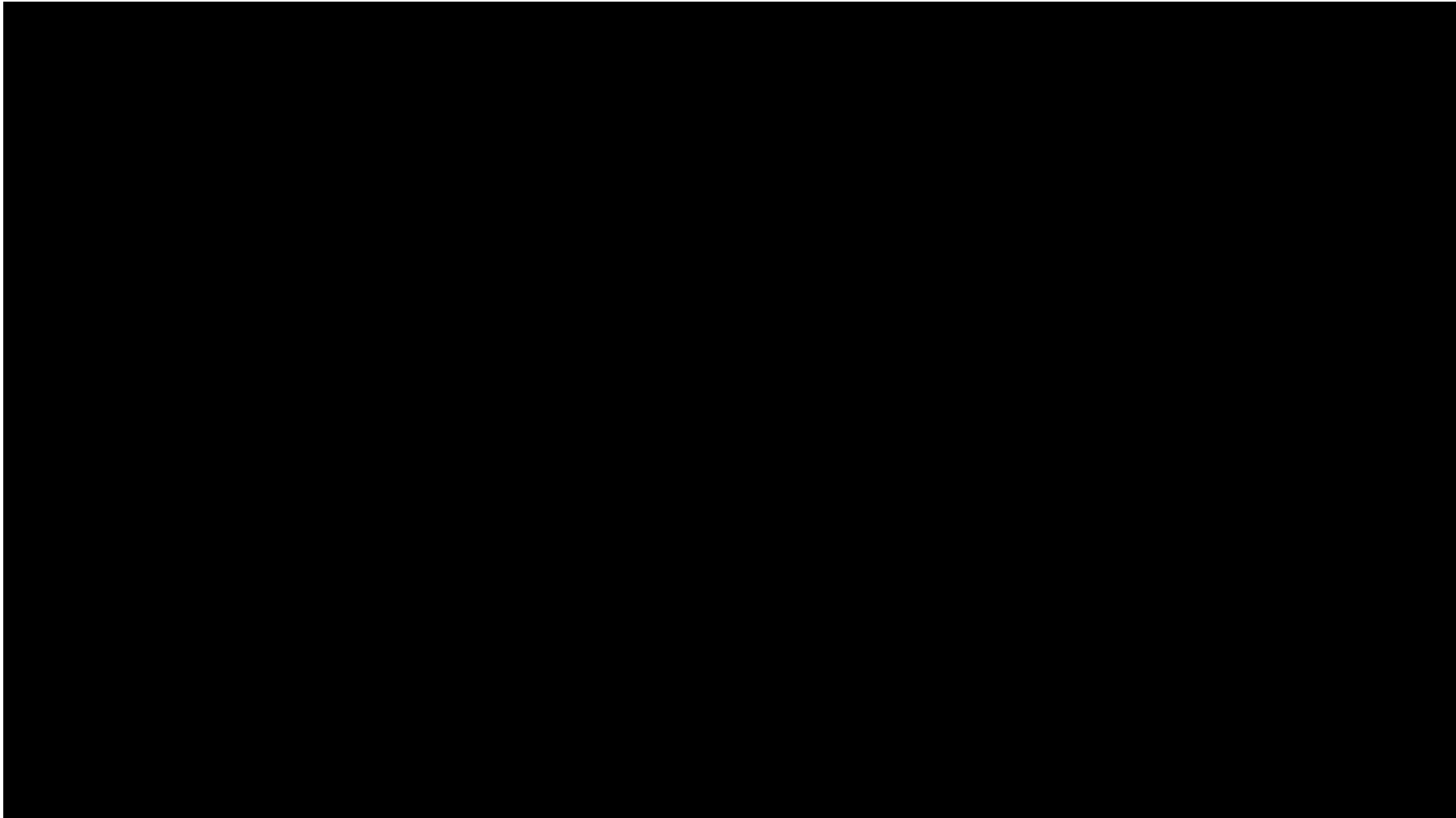
Three Key Ingredients for Success
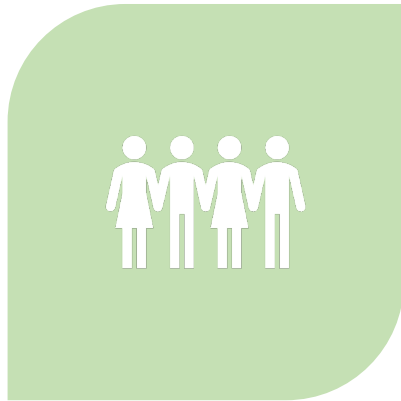
DATA AND DATA
COMMUNITY

INVESTMENT AND
EFFORT

"BIG OPPORTUNITY"

Questions and Discussion