Innovation Opportunities at NCSES

John Finamore, Chief Statistician

October 22, 2024

NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS
NATIONAL SCIENCE FOUNDATION

Outline

- Introduction to NCSES
- Motivation for innovation
- Innovation framework
- Innovation examples
- Final thoughts

Introduction to NCSES

National Center for Science and Engineering Statistics

Measuring America's progress in science, technology, and innovation

Part of the National Science Foundation (NSF)

Located within the Directorate for Social, Behavioral, and Economic Sciences (SBE)

One of 13 principal federal statistical agencies

Overseen by the U.S. Chief Statistician within the White House Office of Management and Budget (OMB)

MISSION

Produce policy relevant, policy neutral **statistical information** on the **U.S. science and engineering (S&E) enterprise**:

- Research and development
- Science and engineering workforce
- U.S. competitiveness in STEM
- STEM education in the United States

Motivation for Innovation

Innovation is our responsibility

*Statistical agencies **should be innovative** in applying new technologies in their methods for designing, collecting, processing, editing, compiling, storing, analyzing, releasing, and disseminating data to improve the accuracy and timeliness of their information and the efficiency of their operations.*

- Statistical Policy Directive #1

Motivation for innovation is clarified by our fundamental responsibilities

The Trust Regulation codifies and clarifies the four fundamental responsibilities of federal statistical agencies:

- Relevance and timeliness
- Credibility and accuracy
- Objectivity
- Confidentiality

NATIONAL ARCHIVES

FEDERAL REGISTER

The Daily Journal of the United States Government

Rule

Fundamental Responsibilities of Recognized Statistical Agencies and Units

A Rule by the Management and Budget Office on 10/11/2024

PUBLISHED DOCUMENT: 2024-23536 (89 FR 82453)

PDF
Document Details
Document Dates
Table of Contents
Public Comments
Regulations.gov Data
Sharing
Print
Document Statistics

DOCUMENT HEADINGS

Office of Management and Budget
5 CFR Part 1321
RIN 0348-AB81

AGENCY:

Office of Management and Budget (OMB), Executive Office of the President.

ACTION:

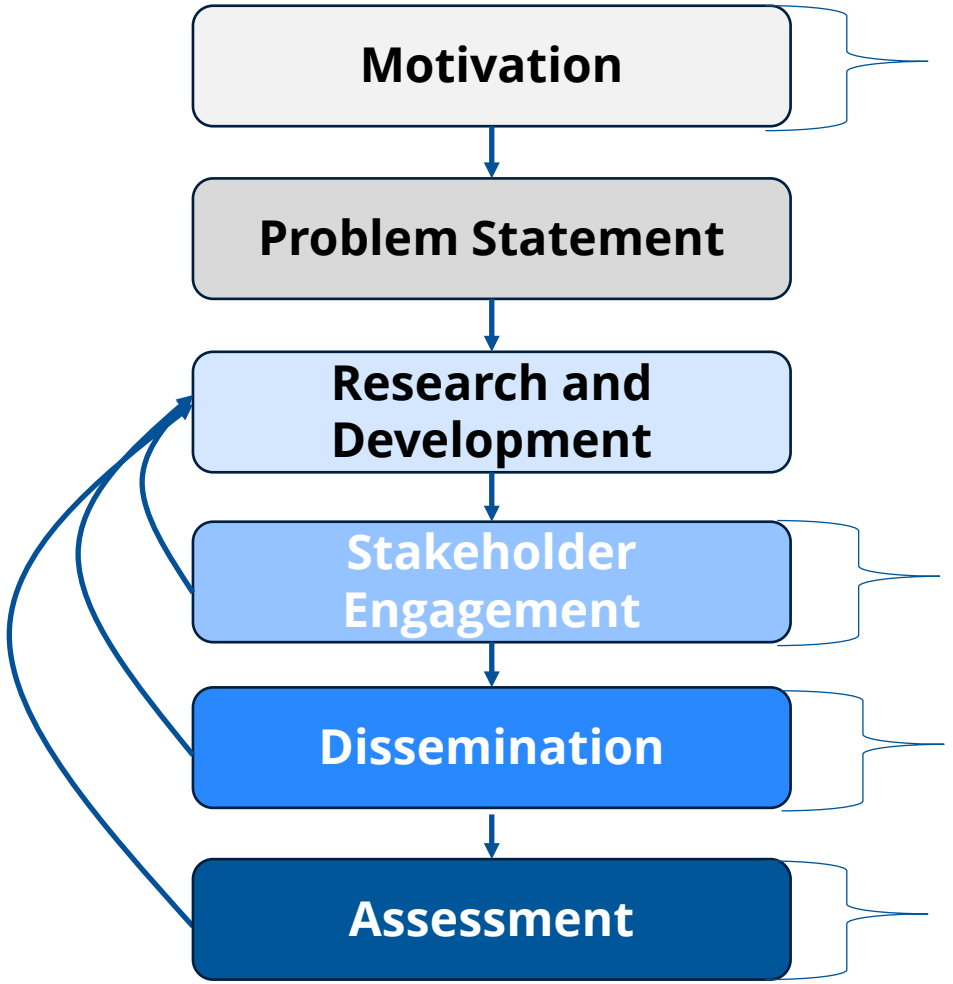
Final rule.

SUMMARY:

Public trust in Federal statistics is essential to their value and use in informing decisions across public and private sectors. To promote public trust in the statistical agencies and units that produce Federal statistics, the Office of Management and Budget issues this final rule pursuant to Title III of the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) to provide direction to Recognized Statistical Agencies and Units (RSAUs) in carrying out their four fundamental responsibilities: produce and disseminate relevant and timely statistical information, conduct credible

Innovation Framework

Innovation Framework



- Fundamental responsibilities of federal statistical agencies

- Collaboration with partner agencies
- Interaction with community (e.g., expert panels; focus groups)

- GovNotice delivery, NCSES website, and social media
- Documentation and user guidance

- User feedback
- Customer satisfaction surveys

Innovation Example:
Survey of Earned Doctorates Restricted Data Access System (SED RDAS)

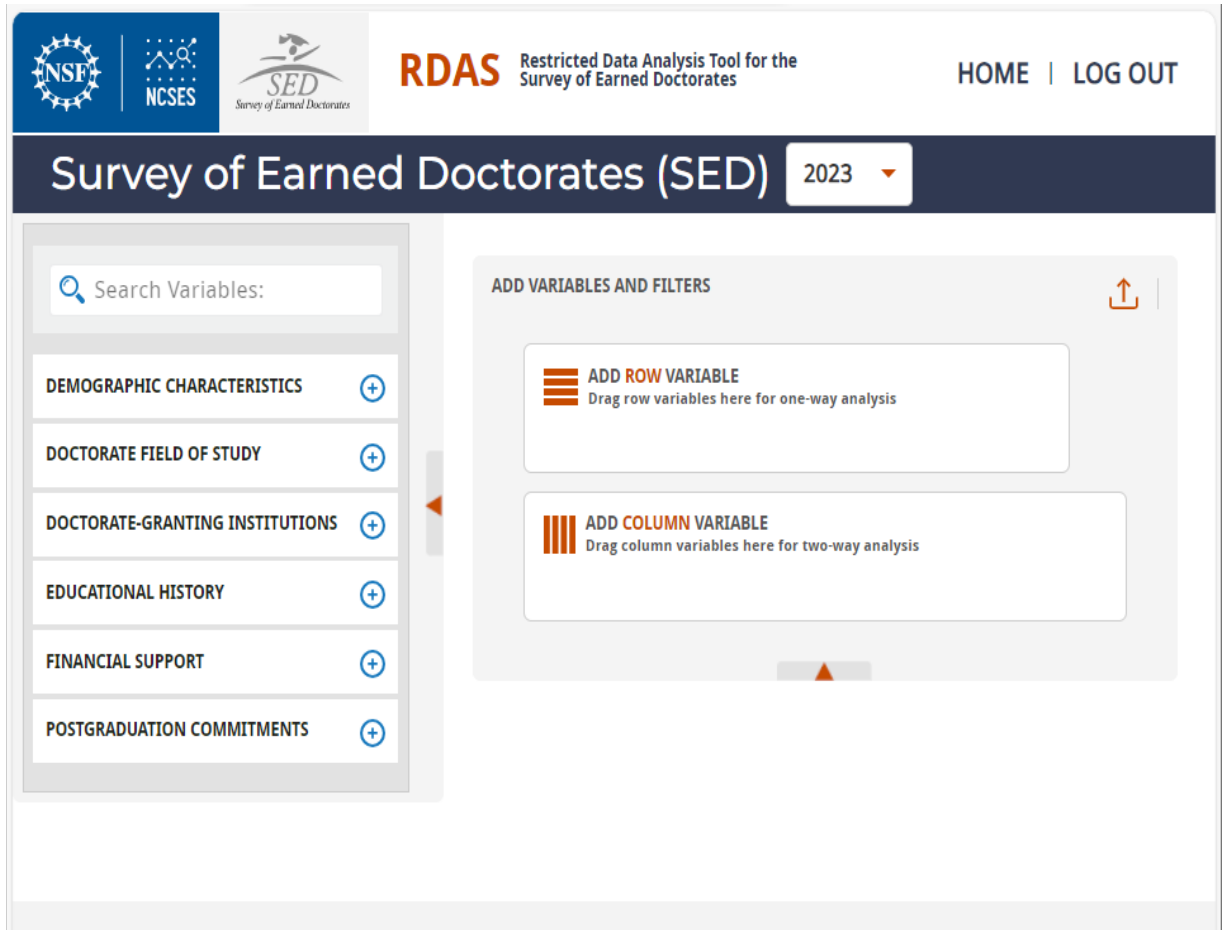
SED RDAS: Background

Survey of Earned Doctorates

- Annual census of all new Research Doctorates Recipients from U.S. academic institutions
- Data collected includes educational history, education-related debt, post graduation plans, demographic information
- Disclosure risk limits publicly available interactive data dissemination options

Motivation	• Relevance and timeliness • Confidentiality
Problem Statement	Users need publicly available interactive SED data access options

SED RDAS: Research and Development



- RDAS is an online tool designed to expand public access to SED data
- Balances utility and privacy by offering access to a subset of variables previously unavailable publicly for a sample of SED respondents (rather than the full set of respondents)
- Estimates differ slightly from official SED estimates, so communication is needed to ensure appropriate usage

Innovation Example:
Data Quality Scorecard

DQ Scorecard: Background

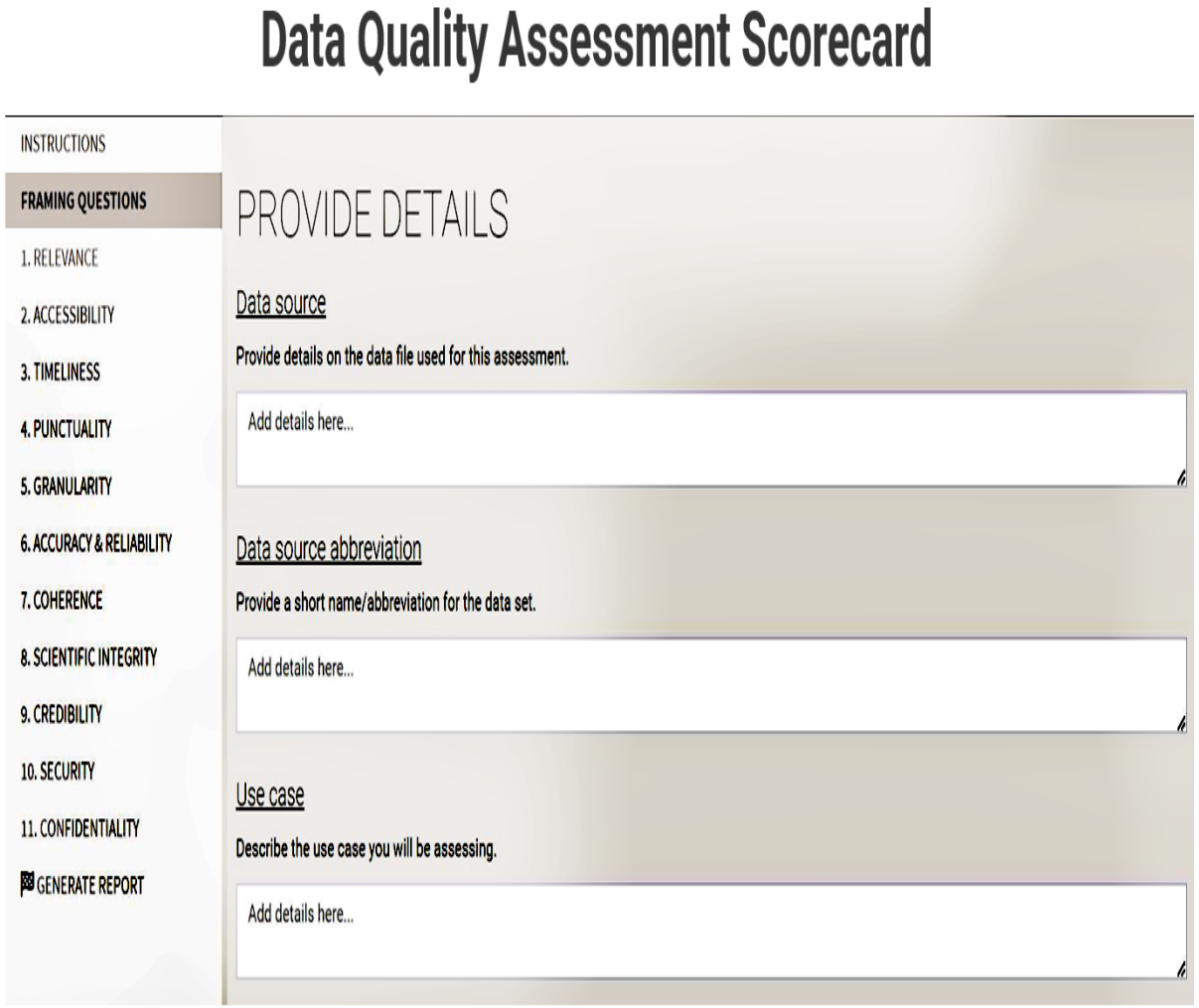
A Variety of Data Sources

- NCSES official statistics
  - Universities
  - Government agencies
  - Businesses
  - Individuals
- Census, BLS, NCES, etc.
- International, OECD data
- Bibliometric/Patent data

- NCSES, in its federal clearinghouse role, uses a variety of data sources to inform its congressionally mandated reports
- New data sources are needed to inform evolving policy needs

Motivation	• Relevance and timeliness • Objectivity
Problem Statement	We need an objective and consistent approach to assess the appropriateness of new data sources to inform policy discussions

DQ Scorecard: Research and Development



- Operationalizes the FCSM Framework for Data Quality
- Enables a fitness for use assessment of a data source for a specific use case
- Informs a broad and consistent understanding of data quality

Innovation Example:
America's DataHub

America's DataHub: Background

- SPD #1 states that statistical agencies **should be innovative**

- Standard procurement completion time **is 208 days**

  (Source: Defense Acquisition University)

Motivation	Relevance and timeliness
Problem Statement	We need an efficient procurement mechanism that supports innovative activities

America's DataHub: Establishment

- Streamlines the procurement process
- America's DataHub is a dynamic consortium of experts, organizations, and ideas that encourages partnerships to address today's most challenging topics
- In two years, NCSES has initiated over 20 projects through America's DataHub

AMERICA'S DATAHUB
CONSORTIUM

Final Thoughts

Final Thoughts

- We encounter opportunities for innovation every day
- Innovation takes resources, time, agency commitment, and patience
- As you're hearing today, all the statistical agencies are exploring innovation (to me, this sounds like a great opportunity for partnerships…)

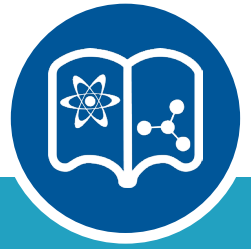
John Finamore

jfinamor@nsf.gov

NCSES's surveys and analysis span the U.S. STEM ecosystem

Research and development

Science and engineering workforce



STEM education

Innovation and competitiveness

NCSES core activities

Gathers data through surveys and other collections to produce official statistics on science and engineering

Performs analysis related to the U.S. S&E enterprise using data from NCSES and other sources

Disseminates the information through public datasets, interactive tools, restricted-use data, and coordinated data infrastructure initiatives

SED RDAS: Engagement, Dissemination, Assessment

Stakeholder Engagement	On-going communication with stakeholder community to refine variable offerings
Dissemination	Emphasis on comprehensive documentation to describe RDAS methodology, functionality, and appropriate usage
Assessment	Disclosure risk assessment is an on-going effort as new SED data products are developed and RDAS variable offerings are refined

RDAS

Restricted Data Analysis System (RDAS)

The Restricted Data Analysis System (RDAS) is an online tool designed to expand public access to the Survey of Earned Doctorates (SED) data. The SED RDAS data is based on a sample subjected to statistical procedures to produce estimates that reflect the entire population of doctorate recipients and protect the confidentiality of individual level data. The RDAS provides a secure platform to analyze a comprehensive set of SED variables from the 2017—2023 SED collections.

Set up an account at ncsesdata.nsf.gov/rdas/

Survey of Earned Doctorates (SED) Data in RDAS

The SED is an annual census of approximately 55,000 individuals receiving a new research doctorate from U.S. colleges and universities. The SED collects information on the doctorate recipient's educational history, funding sources postgraduation plans, employment related and demographic characteristics. The statistical methodology used to produce SED RDAS data protects against disclosure of confidential information, and appropriate adjustments for unit and item nonresponses. To learn more details, see SED RDAS Data Preparation Methodology.

What tables can I create using RDAS?

- One-way counts and percentage distributions
- Multi-way counts and percentage distributions
- Multi-way counts and average/median tables
- Multi-way counts, percents greater/less than, and percents equal/not equal
- Filter by categorical variables
- Apply nested variables

Why should I use RDAS?

- Access the SED restricted data in a secure and on-demand table generation tool
- Create data tables that provide estimates for the full doctorate population without missing data
- Instantly create complex SED data tables in a simple, click-and-drag interface
- Quickly search SED variables and variable information
- Save data tables as CSV and PDF or share table links with other SED data users

DQ Scorecard: Engagement, Dissemination, Assessment

Stakeholder Engagement	On-going communication with stakeholder community to discuss functionality (conference, small group presentations)
Dissemination	Localized dissemination (within NCSES and FCSM)
Assessment	Question wording expert review and considerations of scorecard usage