



Linking the Policy and Utility of Linked Data in Science and Engineering

John Finamore, Chief Statistician

October 26, 2023


NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS
NATIONAL SCIENCE FOUNDATION

Outline

- Introduction to NCSES
 - Linkage Motivation and Goals
 - Current State
 - Vision for the Future
 - Questions for Consideration
-
-
-


Introduction to NCSES

Located within NSF's SBE Directorate, NCSES is one of 13 principal federal statistical agencies

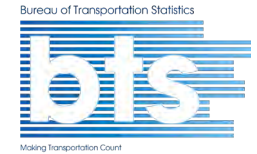


NSF National Science Foundation

Directorate for Social, Behavioral, and Economic Sciences



National Center for Science and Engineering Statistics



NCSES Mandate

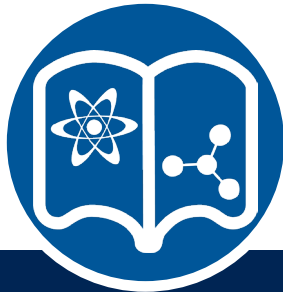
Serve as a central **Federal clearinghouse** for the collection, interpretation, analysis, and dissemination of **objective data** on the **U.S. science and engineering enterprise**

Section 505 of the America COMPETES Reauthorization Act of 2010

NCSES's areas of interest



Science and
engineering
workforce



STEM education



Research and
development



Higher
education
R&D

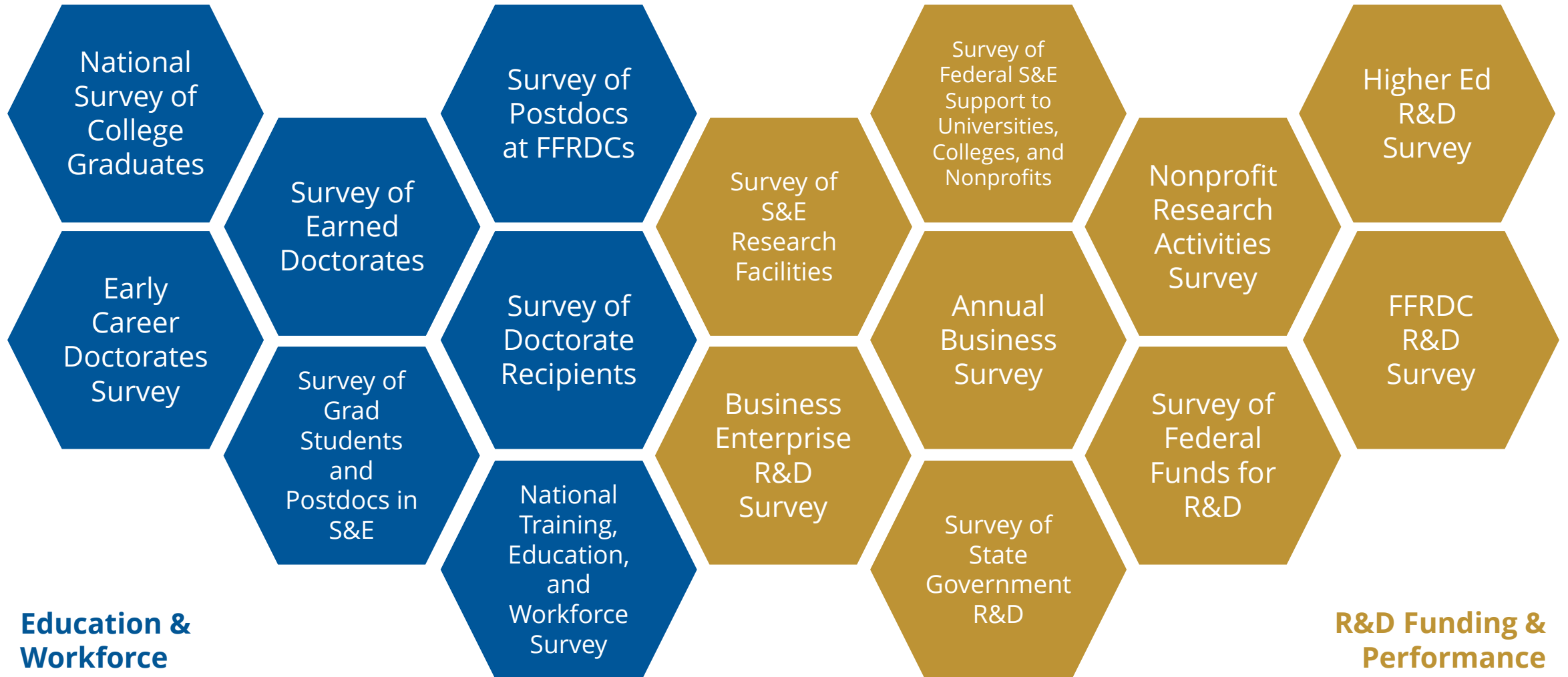


Government
R&D funding

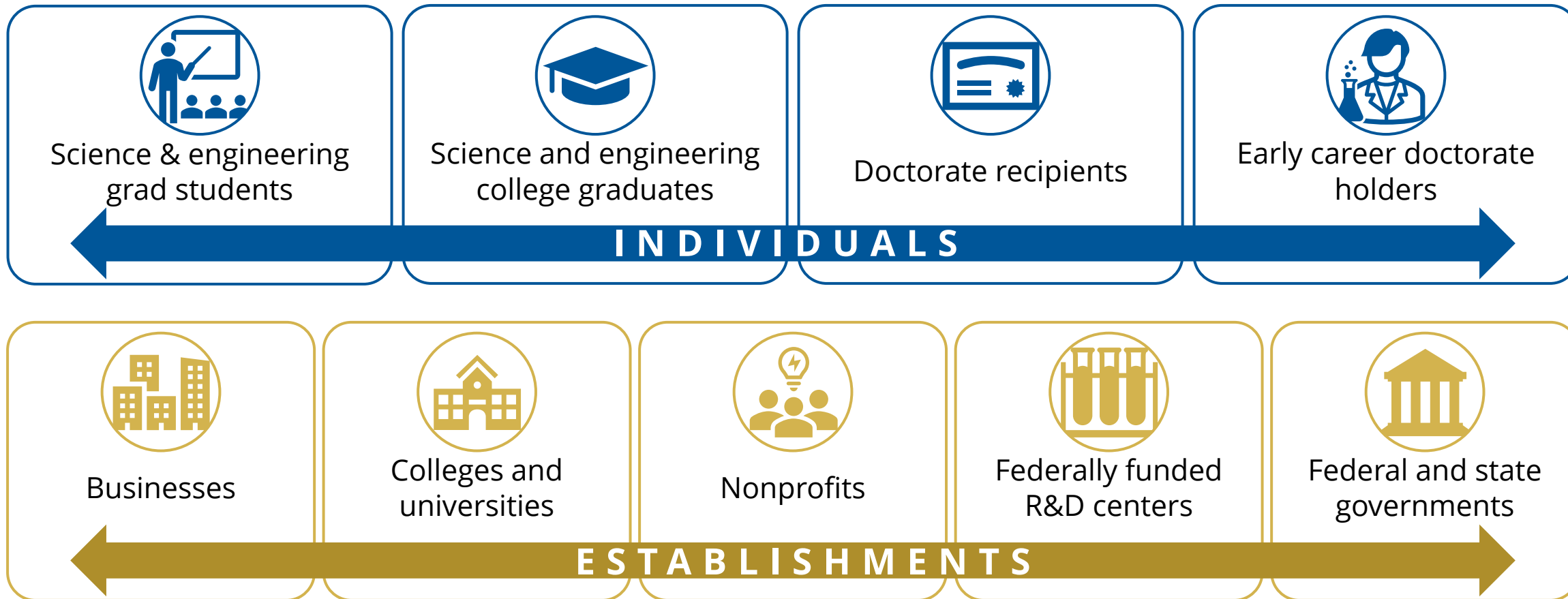


Innovation and
competitiveness

NCSES's survey portfolio



Our surveys collect information from people and organizations



Linkage Motivation and Goals

Linkage Motivation

- We have important research and policy questions
 - Equity
 - Return on investment
 - Prevalence
- Some questions cannot be answered by existing survey data
- Evidence Act encourages the secure use of existing federal data to inform Federal decision-making (e.g., §3581)

Linkage Goals

- Increase the utility of our data products
 - Data quality
 - Survey operations
 - Critical research and policy questions
- Broaden the secure use of NCSES data
 - Linkage program
 - Data access policy
 - Data infrastructure and documentation

Current State

Current approach

- Linkage policy
 - Criteria
 - Quality assessment
 - Secure use of linked data
- Infrastructure development
 - Partitioned space within the NCSES Virtual Data Enclave
 - Data documentation



Current linkages – Survey operations

- National Survey of College Graduates (NSCG)
 - Examines the relationship of degree field and occupation for the nation's college-educated population
 - Longitudinal biennial survey
- Master Address File – Auxiliary Reference File (MAF-ARF)
 - Contains residence information for individuals with an assigned Protected Identification Key (PIK)
- Findings for Respondent Locating
 - MAF-ARF information suggests half of nonrespondents still reside at initial NSCG address
 - Potential to enhance respondent retention approach

Current linkages – Data quality

- National Survey of College Graduates (NSCG)
 - Examines the relationship of degree field and occupation for the nation's college-educated population
 - Longitudinal biennial survey
- Longitudinal Employer-Household Dynamics (LEHD)
 - Quarterly database of linked employer-employee data
 - Tracks aggregate and individual level employment, earnings, and job flows
- Findings
 - Annual earnings – Large differences in linked values among highest workers
 - Employment history – The agreement rates indicate LEHD potentially useful for tracking employment status

Current linkages – Research and policy questions

- Survey of Earned Doctorates (SED)
 - Annual census of research doctorates from U.S. academic institutions
 - Collects information on educational history, demographic characteristics, and postgraduation plans
- UMETRICS (Universities Measuring the Effects of Research on Innovation, Competitiveness, and Science)
 - Transactional data for federally funded grants to institutions
 - Allows the investigation of career outcomes (postdoc), employment placement, and research outcomes (e.g., publications, citations)
- Possible topics of interest
 - Examine grant award rates, career pathways, and research production by demographic characteristics

Current linkages – Continued

- Skilled Technical Workforce

- Examination of various administrative data sources for quality assessment, survey operations, and additional granularity
 - National Student Clearinghouse
 - Participant Individual Record Layout (PIRL)
- National Plan and Provider Enumeration System (NPPES) to supplement or replace credential information

- Research outcomes

- Web of Science (WoS) and Scopus are leading databases providing reference and citation data from academic journals, conference proceedings, and other documents in various academic disciplines



Scopus

Current linkages – Continued even more

- National Secure Data Service Demonstration Project
 - Utilizing PPRL to link data from two statistical agencies (NCHS and NCSES)
 - Utilizing PPRL to link data from with parent agency data and statistical agencies (NSF and NCSES) to inform programs and policies

Vision for the Future

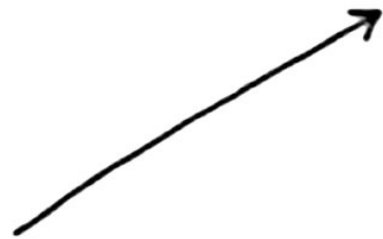
Vision

- Prioritize funding and resources for the establishment of a comprehensive NCSES linkage program
 - Identify key NCSES research and policy questions (e.g., learning agenda)
 - Prioritize linkage opportunities (e.g., STW measurement, PPRL research)
 - Establish a secure compute environment for linkage activities
 - Coordinate data sharing agreements
 - Develop and implement linkage methodology and assessments
 - Establish a data infrastructure (inventory; standardization of variables)
 - Other considerations (e.g., consent, privacy)

Vision Implementation

- It's likely establishing a linkage program will NOT be a straight line

Success



what people think
it looks like

Success



what it really
looks like

Questions for Consideration

Questions for Consideration

- How can the federal statistical system develop guidance or best practices to help agencies establish sustainable linkage programs?
- What linkage challenges have agencies faced and how have they overcome those challenges?





John Finamore

jfinamor@nsf.gov

 <https://ncses.nsf.gov>
 [@NCSESgov](https://twitter.com/NCSESgov)