



Statistics and Tools to Measure the Scientific Enterprise

**Symposium of the U.S. Statistical Agencies
Statistics for Economists
International Year of Statistics**

November 13, 2013

Mr. John R. Gawalt
Director

National Center for Science and Engineering Statistics
www.nsf.gov/statistics



National Center for Science and Engineering Statistics

- **Federal statistical agency within NSF**
- **Responsibility for statistical data on:**
 - Research and development (R&D)
 - S&E workforce
 - US competitiveness in science and technology (S&T) and R&D
 - Condition and progress of science, technology, engineering and mathematics (STEM) education
- **Core Activities:**
 - Collect, acquire, analyze, report and disseminate statistical data
 - Support research using NCSES data
 - Methodological research
 - Education and training of researchers in large-scale nationally representative data sets



Measuring the S&T Enterprise

- **Clearinghouse mandate** - keep track of the health & status of the U.S. science and technology (S&T) enterprise
- **How do we do it?**
 - Conduct surveys
 - Synthesize data and information
 - Promote collection of comparable international data
 - Disseminate information - reports, electronic products, databases, and data files
 - Support fundamental research



Community Partnerships

- **NCSES information consumers**
 - Government/academic/private-sector policymakers
 - Academic and other researchers
 - Nonprofit organizations
 - Professional associations
 - Media
 - General public
- **Communication and outreach**
 - Formal meetings
 - Formal and ad hoc advisory boards
 - External reviews of our work
 - Participation in community events (conferences, meetings, etc.)
 - Responding to information requests
- **Feed the outreach back into NCSES activities**



R&D in the National Economy – Government partnership

- **Issue:** National economic accounting did not separately identify the contribution of R&D and other intangible assets to economic growth.
- **Partnership:** NCSES and the Bureau of Economic Analysis engaged in a multi-year activity to build experimental estimates to demonstrate how business spending on R&D would affect the national accounts and U.S. gross domestic product. R&D would be treated as an investment instead of an expense.
- **Data Source Improvement:** NCSES's Business R&D and Innovation Survey was modified to collect additional and improved R&D statistics for R&D intensive industries.
- **Impact:** GDP would have been higher by an average of 2.9% between 1959 and 2004 if R&D was treated as an investment. BEA has now incorporated this new methodology into the National Income and Product Accounts and GDP measures.



College Graduates in the U.S. – Community partnership

- **Issue:** Census data on college graduates did not include information about their field of study, limiting our understanding of the workforce. It also made NCSES's follow-up of Census respondents who were science and engineering graduates inefficient.
- **Partnership:** NCSES lead an effort, with support from our community, to add a field of degree question to the American Community Survey, which is conducted by the U.S. Census Bureau.
- **Data Source Improvement:** Field of Degree question now appears on Census's ACS.
- **Impact:** 1) NCSES has a more efficient sampling frame for its surveys of the science and engineering workforce, and 2) there are now national estimates of college graduates, by their field of study, in the American Community Survey

S&T Indicators – International partnership

- **Issue:** Placing U.S. measures in the global context is essential for international comparisons. Ensuring that the measures are consistent is key to this goal.
- **Partnership:** NCSES staff serve on a variety of international groups through the Organization for Economic Co-operation and Development to aid in the development of standards to measure R&D and human capital in S&T.
- **Data Source Improvement:** Almost all NCSES surveys have been either the starting point, test bed or expansion point for standard measures on education, workforce or R&D measures for use in international comparisons.
- **Impact:** U.S. concepts, based on NCSES work, have become the benchmark for international data, particularly in business R&D and the highly-skilled workforce



Demographic Differences in STEM – Academic partnership

- **Issue:** Participation in STEM education and the workforce has varied across demographic groups. Accurate measurement of all groups (by age, gender, race/ethnicity, etc.) at different stages of the pipeline can inform our understanding of the economic impact of such participation.
- **Partnership:** NCSES supports the academic community in using our human capital survey data to understanding demographic differences in STEM participation.
- **Data Source Improvement:** Research by the academic community on the sources of demographic differences in STEM have lead to changes in the questions that are asked to the science and engineering community, including the role of dual careers, job satisfaction, and family formation and composition.
- **Impact:** Demographic measures allow us to see the differences among groups – the addition of other qualitative measures allows understanding of why the differences exist.



Ongoing Economic Research

- [R&D Agglomeration](#) (geographic clustering of industrial R&D activity)
- [Race and Gender Variation in STEM Employment and Retention](#)
- [The Impact of R&D Practices on R&D Effectiveness](#)
- [Who Leaves and Who Arrives? Mapping the Connections between Universities and the Science and Engineering Workforce](#)
- [Connecting Outcome Measures of Entrepreneurship, Technology and Science](#)
- [The Long-Term Regional Economic Impacts from Public Investment in University Research](#)
- [The Economic Spillovers from Science](#)



Economic Questions That NCSES Data Addresses

- How much debt to science and engineering students accumulate?
- What are the unemployment rates of college graduates?
- What are sources and levels of funding for R&D activities in the U.S.?
- How has patent and publication activity in the sciences changed over time?
- What is the cost of college attendance?
- How have public attitudes and knowledge about science changed over time?

Data Access – Publications and Products

- **InfoBriefs**
Highlight results from recent surveys and analyses
- **Detailed Statistical Tables (DSTs)**
Contain extensive tables from a particular survey (electronic only)
- **Periodic Overview Reports**
 - *Science and Engineering Indicators*
 - *National Patterns of R&D Resources*
 - *Women, Minorities, and Persons With Disabilities in Science and Engineering*
- **Working papers**
- **Online state and institutional profiles**

Data Access – Online Data Tools/Facilities

- WebCASPAR**

A compilation of databases and a table generator that provides easy access to a large body of data collected from U.S. academic institutions on topics such as degrees awarded, enrollments and R&D expenditures

The screenshot shows the WebCASPAR website interface. At the top, there is a navigation bar with links: HOME, TABLE BUILDER, FIND A VARIABLE, MY WEBCASPAR, DATA UPDATES, TUTORIALS, FAQ, INSTITUTION PROFILES. Below this is a section titled 'Table Builder: create a data table'. It contains two columns of data sources, each with a checkbox and a link to more information. The left column lists NSF data sources, and the right column lists NCES data sources. A 'Select Data Source(s)' button is located below the lists. At the bottom of the page, there is a footer with contact information and a session expiration notice.

National Science Foundation (NSF) Data Sources

- NSF Survey of Earned Doctorates/Doctorate Records File [info](#)
(Years Available:1966-2011)
- NSF Survey of Federal Funds for Research and Development [info](#)
(Years Available:1951-2012)
- NSF Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions [info](#)
(Years Available:1971-2009)
- NSF Survey of Research and Development Expenditures at Universities and Colleges/Higher Education Research and Development Survey [info](#)
(Years Available:1972-2011)
- NSF Survey of Science and Engineering Research Facilities [info](#)
(Years Available:2003-2011)
- NSF-NIH Survey of Graduate Students & Postdoctorates in Science and Engineering [info](#)
(Years Available:1972-2011)

National Center for Education Statistics (NCES) Data Sources

- IPEDS Completions Survey [info](#)
(Years Available:1966-2011)
- IPEDS Completions Survey by Race [info](#)
(Years Available:1977-2011)
- IPEDS Enrollment Survey [info](#)
(Years Available:1967-2011)
- IPEDS Institutional Characteristics Survey Tuition Data [info](#)
(Years Available:1969-2011)
- IPEDS Salaries, Tenure, and Fringe Benefits Survey [info](#)
(Years Available:1971-2012)

Select Data Source(s)

Saved Tables: View predefined tables and tables that you have saved

Frequently Requested Tables:

NCES Degrees Awarded by Degree Level and Field [View](#)

This is a National Science Foundation (NSF) Federal Government computer system. Unauthorized attempts to modify any information stored on this system, defeat or circumvent security features, or use this system for other than its intended purposes are illegal and may result in disciplinary action, criminal prosecution, or both.

Your session will expire after 300 minutes of inactivity. When the session expires, table specifications that have not been saved will be lost.

NCSES Home | NSF Privacy Policy | NSF Web Policies | Data Update Feed | E-mail WebCASPAR | Site Map

Last Updated: July 3, 2013

Data Access – Online Data Tools/Facilities (continued)

- [SESTAT Data Tool](#)

An online table generator for NCSES’s three surveys of the science and engineering workforce (the SESTAT surveys)

SESTAT Data Tool
Scientists and Engineers Statistical Data System

SESTAT Home
Public Use Data Files

Select Survey | Choose Variable(s) | Specify Population (optional) | Select Data Type (optional) | Generate Table

Step 1: Select Survey | Guest User | SESTAT Metadata Explorer | Help&Tutorials | Login

Survey: Integrated Survey Data, SESTAT PUBLIC 2010

Select Survey Data	Select Year	Description
Integrated Survey Data, SESTAT PUBLIC	2010	Integrated Survey Data, SESTAT PUBLIC 2010 contains just over 108,300 records of persons with a science, engineering or S&E-related degree and/or occupation, weighted to represent an estimated 26.9 million persons in the U.S. educated or working as scientists or engineers during the reference week of October 1, 2010. Data from three 2010 surveys (<i>Survey of Doctorate Recipients, National Survey of College Graduates, and the National Survey of Recent College Graduates</i>) were integrated into this combined database to provide information about the employment, educational and demographic characteristics of scientists and engineers in the United States. This public database contains variables that were created to protect the confidentiality of individuals. See SESTAT Metadata Explorer for more details
Survey of Doctorate Recipients, SDR PUBLIC	2008	
International Survey of Doctorate Recipients, ISDR PUBLIC	2006	
Recent College Graduates, NSRCG PUBLIC	2003	
National Survey of College Graduates, NSCG	1999	
	1997	
	1995	
	1993	

Next

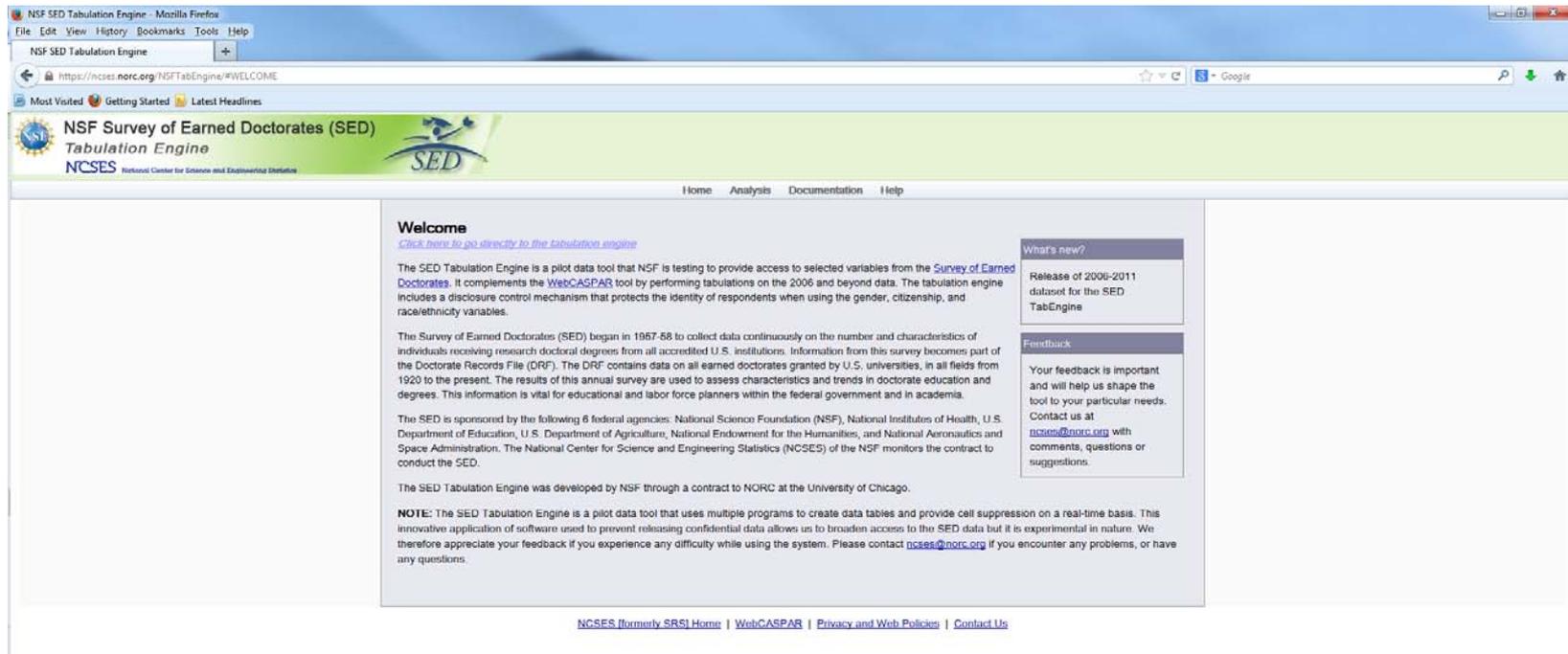
SECURITY: This is a National Science Foundation federal government computer system. Unauthorized attempts to modify any information stored on this system, to defeat or circumvent security features, or to use this system for other than its intended purposes are illegal and may result in disciplinary action, criminal prosecution, or both.

National Center for Science and Engineering Statistics (NCSES)
 The National Science Foundation, 4201 Wilson Boulevard, Arlington, Virginia 22230, USA
 Tel:(703) 292-5111, FIRS:(800) 877-8339 | TDD:(800) 281-8749
 Questions or need help, contact us at sestat@nsf.gov

NSF Privacy Policy
NSF Web Policies

Data Access – Online Data Tools/Facilities (continued)

- **Secure Data Access Facility (SDAF at NORC Data Enclave) –**
 - (1) A virtual data access facility that is currently being built that will provide secure data access to confidential microdata from NCSES surveys for statistical analysis
 - (2) A [table generator](#) to provide access and statistical disclosure limitation procedures to the Survey of Earned Doctorates



The screenshot shows a web browser window titled "NSF SED Tabulation Engine - Mozilla Firefox". The address bar shows the URL "https://nces.norc.org/NSFTabEngine/#WELCOME". The page header includes the NSF Survey of Earned Doctorates (SED) Tabulation Engine logo and navigation links for Home, Analysis, Documentation, and Help.

Welcome
[Click here to go directly to the tabulation engine](#)

The SED Tabulation Engine is a pilot data tool that NSF is testing to provide access to selected variables from the [Survey of Earned Doctorates](#). It complements the [WebCASPAR](#) tool by performing tabulations on the 2006 and beyond data. The tabulation engine includes a disclosure control mechanism that protects the identity of respondents when using the gender, citizenship, and race/ethnicity variables.

The Survey of Earned Doctorates (SED) began in 1967-68 to collect data continuously on the number and characteristics of individuals receiving research doctoral degrees from all accredited U.S. institutions. Information from this survey becomes part of the Doctorate Records File (DRF). The DRF contains data on all earned doctorates granted by U.S. universities, in all fields from 1920 to the present. The results of this annual survey are used to assess characteristics and trends in doctorate education and degrees. This information is vital for educational and labor force planners within the federal government and in academia.

The SED is sponsored by the following 6 federal agencies: National Science Foundation (NSF), National Institutes of Health, U.S. Department of Education, U.S. Department of Agriculture, National Endowment for the Humanities, and National Aeronautics and Space Administration. The National Center for Science and Engineering Statistics (NCSES) of the NSF monitors the contract to conduct the SED.

The SED Tabulation Engine was developed by NSF through a contract to NORC at the University of Chicago.

NOTE: The SED Tabulation Engine is a pilot data tool that uses multiple programs to create data tables and provide cell suppression on a real-time basis. This innovative application of software used to prevent releasing confidential data allows us to broaden access to the SED data but it is experimental in nature. We therefore appreciate your feedback if you experience any difficulty while using the system. Please contact nces@norc.org if you encounter any problems, or have any questions.

What's new?
 Release of 2006-2011 dataset for the SED TabEngine

Feedback
 Your feedback is important and will help us shape the tool to your particular needs. Contact us at nces@norc.org with comments, questions or suggestions.

At the bottom of the page, there are links for [NCSES \(formerly SRS\) Home](#), [WebCASPAR](#), [Privacy and Web Policies](#), and [Contact Us](#).

Data Access – Online Data Tools/Facilities (continued)

- Census Research Data Centers

To access data from NCSES's Business R&D and Innovation Survey

The screenshot shows the website for the Center for Economic Studies (CES) at the Census Bureau. The page is titled "Center for Economic Studies (CES)" and features a navigation menu with options like "People", "Business", "Geography", "Data", "Research", and "Newsroom".

RDC Research Opportunities

Qualified researchers with approved Research Data Center (RDC) projects

- Gain restricted access to the data at secure [Census Bureau Research Data Center \(RDC\) locations](#)
- Are sworn for life to [protect the confidentiality](#) of the data they access
- Provide benefits to programs of the data-providing agency or agencies
- Publish findings from approved projects on a wide variety of social science topics in peer-reviewed journals

Proposals

Projects using Census Bureau data must...

- Provide benefit to Census Bureau programs
- Demonstrate scientific merit
- Require non-public data
- Be feasible given the data
- Pose no risk of disclosure

More Information

- [Role of RDCs](#)
- [Data](#)
- [Locations](#)
- [Secure RDC Environment](#)
- [How to Apply](#)

For projects using data from partnering agencies, see:

- [Agency for Healthcare Research and Quality \(AHRQ\)](#)
- [National Center for Health Statistics \(NCHS\)](#)

Research Data Centers

A map of the United States shows the locations of Research Data Centers. A list of locations includes:

- Atlanta
- Boston
- California, Berkeley
- California, Irvine (Coring Room)
- California, Los Angeles
- California, San Diego
- California, USC (Coring Room)
- Census Headquarters
- Chicago
- Michigan
- Minnesota
- New York, Baruch
- New York, Cornell
- Northwest
- Penn State (Coring Room)
- Texas
- Texas, LISA
- Triangle, RTI International

Interested in hosting a Research Data Center at your institution?

Please contact our [Director of Research](#) for information on physical security requirements, IT infrastructure costs, on-site staffing requirements, and procedures for opening a Research Data Center.

Data Access – Microdata

- Public-Use Files (PUFs)

- Microdata files with confidentiality protections applied (if necessary)
- Downloadable microdata files for some surveys; microdata access through WebCASPAR for others

- Restricted-Use Files (RUFs)

- Microdata files with limited confidentiality protections applied
- Selected NCSES files available under a licensing agreement
- Licensees physically hold the microdata
- Business R&D and Innovation Survey available through the Census RDC's



Research Support Mechanisms

- **NCSES grants program**

[Research on the S&T Enterprise: Statistics and Surveys](#)

- **Grants programs in NSF/SBE**

[Science of Science and Innovation Policy \(SciSIP\)](#)

[Economics](#)

[Methodology, Measurement and Statistics \(MMS\)](#)

[Science of Organizations \(SoO\)](#)

[Science, Technology and Society \(STS\)](#)

[Other Opportunities in the Social, Behavioral and Economic Sciences](#)

- **NSF-sponsored Research Opportunities**

[American Educational Research Association](#)

[Association for Institutional Research](#)



Thank You

John R. Gawalt
jgawalt@nsf.gov