

Guide for Public Use Data Files

National Science Foundation's
Higher Education Research and Development Survey:

Fiscal Year 2018

September 2019

National Science Foundation
National Center for Science and Engineering Statistics

Table of Contents

1. HERD SURVEY (FY 2010 – PRESENT)	3
1.1. SURVEY POPULATION, INSTRUMENT, AND DATA AVAILABILITY	4
1.1.1. Target Population and Frame	4
1.1.2. Key Variables.....	5
1.1.3. Changes to the Survey for FY 2018.....	6
1.1.4. Excluded Data.....	7
1.2. DATA IMPUTATION	7
1.2.1. Retroimputation of Prior Years’ Data.....	7
1.3. PUBLIC USE DATA FILES.....	8
1.3.1. Institution Header Records (Short Form and Standard Survey Population Data Files).....	8
1.3.2. Data Records.....	10
2. ACADEMIC R&D SURVEY (FY 1979 – FY 2009)	13
2.1. SURVEY POPULATION, INSTRUMENT, AND DATA AVAILABILITY	13
2.1.1. Population Size and Structure.....	13
2.1.2. Survey Instrument.....	14
2.1.3. Historical Changes to the Survey Instrument	15
2.1.4. FY 2009 Pilot Survey.....	16
2.1.5. Excluded Data.....	16
2.1.6. Data Element Availability by Year.....	16
2.2. DATA IMPUTATION	23
2.2.1. Retroimputation of Prior Years’ Data.....	23
2.3. PUBLIC USE DATA FILES.....	23
2.3.1. Institution Header Records	24
2.3.2. Data Records.....	25
3. ADDITIONAL RESOURCES	27
3.1. Publications	27

3.2. Electronic Access	27
3.3. Contact for More Information	27
4. OTHER FORMATS	28
4.1. HERD (FY 2010 – Present).....	28
4.1.1. Excel Data File Format	28
4.1.2. SAS Data File Format	32
4.1.3. SPSS Data File Format	32
4.1.4. CSV Data File Format	32
4.2 Academic R&D Survey (FY 1979 – FY 2009).....	32
4.2.1 Excel Data File Format	32
4.2.2. SAS Data File Format	34
4.2.3. SPSS Data File Format	34
4.2.4. CSV Data File Format	34

The first section of this data users' guide provides a detailed description of the public use data files for the FY 2018 Higher Education Research and Development (HERD) Survey. The HERD Survey is sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation. The HERD Survey replaces the previous annual collection, the Survey of Research and Development Expenditures at Universities and Colleges (Academic R&D Expenditures Survey), which was conducted from FY 1972 through FY 2009. The second section of this guide provides a detailed description of the public use data files for this previous version of the R&D survey. Each section includes an overview of the surveys and a detailed description of each data file structure. For additional information about the survey population and procedures, refer to the FY 2018 Survey Methodology Report and methodology reports from previous collection cycles (available upon request from the NCSES project manager listed in section 3.3).

1. HERD SURVEY (FY 2010 – Present)

In 2007, NCSES began a large-scale project to redesign and update the Academic R&D Expenditures Survey. One of the goals of this redesign was to minimize the sources of measurement error for the survey. As part of the redesign effort, NCSES held data user workshops and expert panel meetings, worked with accounting and survey methodology experts, and visited more than 40 institutions to receive input on possible changes to the survey. In FY 2009 40 institutions included in the current survey population participated in a pilot test of the newly designed HERD Survey. Beginning with FY 2010 the HERD Survey was conducted with the entire survey population.

With the introduction of the HERD Survey in FY 2010, a number of changes were made to the questionnaire. The HERD Survey continued to ask for current fund expenditures for separately budgeted R&D, but total R&D expenditures now included both S&E fields and non-S&E fields such as the humanities, education, law, and the arts. Although there was one item in the previous survey concerning non-S&E expenditures, most items only requested information on S&E expenditures. All HERD Survey questions asked about R&D across all fields. Additionally, clinical trials and research training grants were explicitly included in the definition of R&D.

Questions 1, 7–9, 11, and 14 are very similar to items found in the Academic R&D Expenditures Survey; however, all questions were reformatted in some way as part of the survey redesign. For example, in past years respondents were asked to report federal and total expenditures in a certain category. In the HERD Survey they were asked to report both federal and nonfederal expenditures. Questions 2–6, 10, 12, 13, and 15 were not included in the previous survey. In FY 2011 NCSES reformatted the survey by removing question 5.1 and expanding questions 7 and 8 to collect details on businesses and nonprofit organizations. In FY 2012 question 1.1 was added to the survey and requested information on the types of expenditures included in data reported as institutionally financed research (question 1, row e1). In FY 2012 NCSES introduced a shorter version of the HERD Survey questionnaire for all institutions reporting less than \$1 million in total R&D in their FY 2011 report or during the FY 2012 population review. This short form questionnaire requested R&D expenditures by source of funds, information on expenditures reported as institutionally financed organized research, federally and nonfederally financed R&D expenditures by major R&D field, and expenditures received from or passed to higher education institutions. In FY 2015 the question asking for information about federally financed R&D

expenditures from ARRA funds was removed and all subsequent questions were renumbered. In FY 2016 Question 2 was expanded to request foreign funded R&D expenditures by foreign source and the question asking for a headcount of postdocs was removed. In addition, in FY 2016 revisions were made to the fields of R&D included in Questions 9, 11, and 14 of the survey. The revised fields better reflected the R&D currently being conducted at universities and colleges and also make HERD fields more consistent with those used by other NCSSES surveys as well as CIP codes. Additional information about this change can be found in Appendix A.

In addition to these changes in the data collection instrument, changes were also made to the target population and sampling frame. The FY 2009 Academic R&D Expenditures Survey population included all universities and colleges that granted a bachelor's or higher degree in science or engineering and expended \$150,000 or more in S&E R&D in the previous fiscal year. Doctorate-granting institutions and historically black colleges or universities (HBCUs) that did not meet the \$150,000 threshold were excluded from the population. NCSSES changed the target population and sampling frame starting with the FY 2010 collection cycle. The HERD Survey target population included all institutions that granted a bachelor's or higher degree in any academic discipline, had at least \$150,000 in separately budgeted R&D expenditures during the targeted fiscal year, and were geographically separate campuses headed by a president or chancellor.

1.1. SURVEY POPULATION, INSTRUMENT, AND DATA AVAILABILITY

1.1.1. Target Population and Frame

The target population for the FY 2018 HERD Survey included all institutions of higher education in the United States, Guam, Puerto Rico, and the U.S. Virgin Islands that granted a bachelor's or higher degree in any academic discipline and had at least \$150,000 in separately budgeted R&D expenditures during the institution's 2018 fiscal year.

In the years prior to beginning the new survey collection, NCSSES contacted institutions that met the degree-granting criterion but were not in the population of the previous census to determine whether they met the \$150,000 expenditure criterion. For the FY 2010 survey the review of the population of 4-year degree-granting institutions was conducted concurrently to the HERD Survey collection. For this reason the population of the FY 2010 HERD Survey was comprised only of institutions that were part of the FY 2009 Academic R&D Expenditures Survey. Beginning in FY 2011 the review of the population of 4-year degree-granting institutions was conducted prior to each collection.

In FY 2010 NCSSES revised the reporting procedures to require survey responses from each geographically separate institution campus headed by a president or chancellor. The goals of the change to the reporting unit were to improve the comparability of institutions appearing in the ranking tables and to facilitate peer analysis by having data available at a similar level of detail for all institutions. As a result of this step, the overall number of academic institutions in the population increased from 711 in FY 2009 to 744 in FY 2010. In FY 2018 the population increased from 898 in FY 2017 to 905 institutions. Of the academic institutions included in the FY 2018 population, 269 were asked to complete the short form survey questionnaire, and 646 completed the full survey questionnaire.

The survey population was reviewed before data collection began to ensure that each institutional classification was accurate. Characteristics of the schools were reviewed before and during the course of the survey to determine whether changes had occurred (e.g., highest degree granted; school openings, closings, or mergers).

1.1.2. Key Variables

The FY 2018 HERD Standard Form Survey consisted of the following 16 questions:

- **Question 1** requested that institutions report their total expenditures for separately budgeted R&D by source of funds (e.g., federal, business, nonprofit organizations). **Question 1.1** requested information on the types of expenditures included in the data reported as institutionally financed research in question 1, row e1 (e.g., competitively awarded internal grants for research, startup packages, bridge funding, seed funding, tuition assistance for student research personnel).
- **Question 2** asked how much of the total R&D expenditures came from foreign sources.
- **Question 3** asked how much of the total R&D expenditures that were externally funded came from contract and grant agreements.
- **Question 4** asked how much of the total R&D expenditures was expended for R&D projects in the institution's medical school.
- **Question 5** asked how much of the total R&D expenditures was expended for clinical trial R&D.
- **Question 6** requested that institutions report the amount of their federal and nonfederal R&D expenditures for basic research, applied research, and development.
- **Question 7** requested that institutions report the amount of their federal and nonfederal R&D expenditures from Question 1 that was received by the institution as a subrecipient from higher education institutions, businesses, nonprofit organizations, and other sources.
- **Question 8** requested that institutions report the amount of their federal and nonfederal R&D expenditures from Question 1 that was passed through to a higher education institution, business, nonprofit organization, or other subrecipient.
- **Question 9A–K** requested that institutions report the amount of their R&D expenditures by field that was funded by federal agency sources (i.e., the U.S. Departments of Agriculture, Defense, Energy, and Health and Human Services; the National Aeronautics and Space Administration (NASA); NSF; and all other federal agencies).
- **Question 10** requested source and R&D expenditures information about agencies that were included in the Other federal agency column from Question 9.

- **Question 11A–K** requested that institutions report the amount of their R&D expenditures by field that were funded by nonfederal agency sources (i.e., state and local governments, businesses, nonprofit organizations, institutional funds, and other).
- **Question 12** requested that institutions report their total R&D expenditures by type of direct and indirect cost.
- **Question 13** requested that institutions report their capitalization thresholds for software and equipment.
- **Question 14A–K** requested that institutions report the portion of their federal and nonfederal R&D expenditures by field that went toward the purchase of capitalized R&D equipment.
- **Question 15** requested that institutions report the headcount of personnel who received the R&D salaries, wages, and fringe benefits reported in Question 13.
- **Question 16** requested that institutions provide contact information for the individual responsible for the survey and one alternative contact, what month their fiscal year ended, and how long it took to complete the survey. Space for additional comments was also included.

The FY 2018 HERD Short Form Survey consisted of the following 5 questions:

- **Question 1** requested that institutions report their total expenditures for separately budgeted R&D by source of funds (e.g., federal, business, nonprofit organizations). **Question 1.1** requested information on the types of expenditures included in the data reported as institutionally financed research in question 1, row e1 (e.g., competitively awarded internal grants for research, startup packages, bridge funding, seed funding, tuition assistance for student research personnel).
- **Question 2** requested that institutions report their federal and nonfederal R&D expenditures by major R&D field.
- **Question 3** requested that institutions report the amount of their federal and nonfederal R&D expenditures from question 1 that was received by the institution as a subrecipient from higher education institutions.
- **Question 4** requested that institutions report the amount of their federal and nonfederal R&D expenditures from question 1 that was passed through to a higher education institution.
- **Question 5** requested that institutions provide contact information for the **individual** responsible for the survey and one alternative contact, what month their fiscal year ended, and how long it took to complete the survey. Space for additional comments was also included.

1.1.3. Changes to the Survey for FY 2018

The survey instructions were revised to clarify that expenditures for institution research administration and support should be excluded.

1.1.4. Excluded Data

Some data collected as part of the HERD Survey are not included in the public use data files. Information categorized as confidential is excluded from publicly available data. Information received from individual institutions regarding the distinction between separately budgeted institutional funding, cost sharing, unrecovered indirect costs, and recovered indirect costs is not published or released; only aggregate totals appear in tabulations.

1.2. DATA IMPUTATION

In order to provide totals of all academic R&D expenditures, it was necessary to develop estimates for the institutions that did not respond to the survey. Data imputation is an automated procedure used to estimate data for totally and partially nonrespondent institutions. Missing values were imputed based on the previous year's data and the reported data of peer institutions in the current cycle.

For institutions that had not responded by the closing date of the survey and had been included in the FY 2017 HERD survey population, R&D expenditures were imputed by applying inflator/deflator factors to the prior year's key totals. The key totals for FY 2018 included total R&D expenditures, federal R&D expenditures, expenditures received as a subrecipient from higher education sources, expenditures received as a subrecipient from non-higher education sources, expenditures passed through to higher education entities, expenditures passed through to non-higher education entities, and total capitalized equipment R&D expenditures. Imputation factors were derived from the data of responding institutions with similar characteristics, including highest degree granted, type of institutional control (public or private), and level of total R&D expenditures. The key variables were then distributed among the various subtotal and detail fields using the same relative percentages that were last reported by that institution. If an institution was imputed in the previous year, the inflator/deflator factors were applied to the imputed data, and the key variables were distributed as described above.

For institutions that were new to the survey population, there were no past year's data to reference. For these institutions, total R&D expenditures were imputed with minimum placeholder values of \$150,000 or \$1 million, depending on the institution's response to the population review screener. Other values were then imputed as a proportion of total R&D expenditures based on the data of institutions with similar characteristics. Data for partial nonresponses were imputed using similar techniques.

Capitalization thresholds for software and equipment were not imputed. (See the imputation procedures in appendix I of the FY 2018 Methodology Report for more information.)

1.2.1. Retroimputation of Prior Years' Data

The last step in the imputation process was to perform a backcasting, or retro-imputation, of previous years' imputed data. If an institution reported expenditures after 1 year or more of

nonresponse, the current year’s data were used to retro-impute previous years’ data. For each institution, formerly imputed key variables that were part of the HERD survey were recomputed to ensure that the imputed data accurately represented the growth patterns shown by data reported on the FY 2018 HERD survey. If, for example, data were reported for FY 2014 and FY 2018 but not for the intervening years, the difference between the reported figures for each item total was calculated and evenly distributed across the intervening years (FYs 2015–17). The new figures were spread across disciplines or sources of support on the basis of the most recent reporting pattern. These procedures result in much more consistent reporting trends for individual institutions but have little effect on aggregate figures reflecting national totals. Retro-imputation completed after the FY 2018 collection altered aggregated values for FY 2017 total R&D expenditures and expenditures from each source of funding by .01% or less. For variables that had never been reported, such as for an institution new to the population in FY 2017 that did not respond until FY 2018, the ratio of the variable to a key value was used to re-impute last year’s data.

1.3. PUBLIC USE DATA FILES

After completion of the annual data processing cycle, the collected data are stored in files for public use and archival purposes. These files include any corrections to prior-year data that survey respondents have submitted. For each survey year, the data file contains two record types: institution header records and data records. Apart from alphabetic information (e.g., names of institutions, cities), most identifying information in the survey data files is represented by numeric codes. The allowed values, their meanings, and file formats are described below.

For FY 2012 to FY 2018 data from institutions that completed the short form version of the survey are published separately from those that completed the standard survey form; for this reason the two populations are treated separately in these data files as well.

1.3.1. Institution Header Records (Short Form and Standard Survey Population Data Files)

The institution header record provides basic identifying data for a given institution. One record of this type exists for each institution for each survey cycle in which the institution participated; this record is always the first record for an institution in a given year. The format is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–10	Fiscal Year
11	Record Type 1= Institution header record

12–19	NCSES Institution ID
20–25	IPEDS Institution ID
26	HBCU Indicator 0 = not HBCU 1 = HBCU Blank = No indicator assigned; aggregation of institutions
27	Medical School Indicator F = Does not have a medical school N = Null; information was not included T = Has a medical school
28	High Hispanic Enrollment Indicator 0 = Is not a high Hispanic enrollment institution 1 = Is a high Hispanic enrollment institution Blank = No indicator assigned; aggregation of institutions
29	Type of Institution 1 = Academic
30	Highest Degree Granted by Institution 1 = Doctorate 2 = Master's 3 = Bachelor's 4 = Associate's 5 = No degree 6 = Professional degree
31	Type of Control 1 = Public 2 = Private ? = No institutional control assigned; aggregation of institutions
32–133	Institution Name
134–153	City Name
154–155	State Abbreviation ?? = No state; aggregation of institutions
156–165	ZIP Code ???? = No ZIP Code; aggregation of institutions
166–192	Not Used

1.3.2. Data Records

Data records contain the numerical responses to the HERD Survey questionnaire items. These records are identified by question and row numbers corresponding to those on the FY 2018 questionnaire. It should be noted that although the meaning of some question and row numbers changed during the period from FY 2010 to FY 2018, the line references for these years have been made consistent with the FY 2018 format to facilitate data trend analyses. All expenditure data items are expressed in thousands of dollars. Data for survey questions 15 and 16 are headcounts and are expressed as actual numbers. A data record is present for each data line on the questionnaire to which a nonzero response has been received. For total rows in each survey question, a data record is present regardless of whether a zero or nonzero response has been received. A status code associated with each data column indicates the source of or special treatment for the data.

The format for the data reported by institutions responding to the standard survey form is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–10	Fiscal Year
11	Record Type 2 = Data record
12	Blank space
13–17	Survey Questionnaire Number
18	Blank space
19–108	Survey Question Content Description
109	Blank space
110–199	Survey Question Row Description
200	Blank space
201–230	Survey Question Column Description
231	Blank space
232–239	Data Question 1.1 Response Codes -1=Don't Know

0=No
1=Yes

- 240 Status Code for Data
Blank = Normal Response
i = Imputed by Computer for Nonresponse
n = Data Not Available
- 241–390 Other Information
Reason Specific Types of Funding listed in Question 1.1 were not
Included in Data Reported for Current Year (FY 2012 and later years)
Specific Federal Agencies Listed in Survey Question 10
Inclusion of Clinical Trials in FY 2009 Submission for Question 5.1 (FY 2010
files only)
- 391 Status Code for Other Information
Blank = Normal Response
i = Imputed by Computer for Nonresponse
n = Data Not Available
- 392-441 Standardized Name for Specific Federal Agency Listed in Question 10

The format for the data reported by institutions responding to the survey short form is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–10	Fiscal Year
11	Record Type 2 = Data record
12	Blank space
13–16	Survey Questionnaire Number
17	Blank space
18–72	Survey Question Description
73	Blank space
74–138	Survey Question Row Description

139	Blank space
140–169	Survey Question Column Description
170	Blank space
171–178	Data Question 1.1 Response Codes -1=Don't Know 0=No 1=Yes
179	Status Code for Data Blank = Normal Response i = Imputed by Computer for Nonresponse n = Data Not Available
180–195	Other Information Reason Specific Types of Funding listed in Question 1.1 were not Included in Data Reported for Current Year (FY 2012 and later years)

2. ACADEMIC R&D SURVEY (FY 1979 – FY 2009)

2.1. SURVEY POPULATION, INSTRUMENT, AND DATA AVAILABILITY

2.1.1. Population Size and Structure

The Academic R&D Expenditures Survey was conducted annually from 1972 through 2009. From FY 1984 through FY 1997 the academic survey was conducted as a population survey every 5 years and as a sample survey in each of the 4 intervening years. Beginning in FY 1998, the survey was conducted as a full population survey every year. Prior to FY 1998 the academic survey population consisted of all universities and colleges that had doctoral programs in science or engineering fields or performed at least \$50,000 in separately budgeted S&E R&D in a given year. The population criteria were revised in FY 1998 to include all universities and colleges that granted a bachelor’s degree or higher in science or engineering and performed at least \$150,000 in separately budgeted R&D in a given fiscal year, all S&E doctorate-granting institutions, and all HBCUs. From FY 2004 through FY 2009, the academic survey population included all universities and colleges that granted a bachelor’s or higher degree in science or engineering and expended \$150,000 or more in S&E R&D in a given year. Doctorate-granting institutions and HBCUs that did not meet the \$150,000 threshold were excluded from the population.

The number of academic institutions included annually in the survey population from 1972 through 2009 is as follows:

FY	Number of Academic Institutions Surveyed	Number of Academic Institutions in File
1972	591	584
1973	591	584
1974	597	594
1975	537	534
1976	536	534
1977	537	535
1978 ⁺	317	315
1979	562	561
1980	560	559
1981	557	557
1982	556	556
1983	560	559
1984 [*]	403	556
1985 [*]	403	556
1986 [*]	403	556
1987 [*]	399	555
1988	556	555
1989 [*]	459	573
1990 [*]	459	573
1991 [*]	459	572

FY	Number of Academic Institutions Surveyed	Number of Academic Institutions in File
1992*	461	575
1993	681	681
1994*	500	681
1995*	499	680
1996*	493	692
1997*	493	692
1998	556	556
1999	598	598
2000	624	624
2001	610	610
2002	626	626
2003	631	631
2004	612	612
2005	640	640
2006	650	650
2007	670	670
2008	690	690
2009	711	711

⁺ This year included surveyed doctorate-granting institutions only.

* This was a sample year.

For additional details about changes to the population and frame, refer to previous year methodology reports.

2.1.2. Survey Instrument

The FY 2009 Academic R&D Expenditures Survey consisted of eight items requesting information on expenditures and awards for R&D activities during the institutions' 2009 fiscal year. Six of the survey items asked for expenditures from current funds for separately budgeted R&D projects in S&E fields. Item 2A asked for R&D expenditures in non-S&E fields, and Item 4 asked for the number and amount of R&D awards that the institution received in both S&E and non-S&E fields.

The survey consisted of eight main items:

- **Item 1** requested that institutions report their total expenditures for separately budgeted S&E R&D by source of funds (e.g., federal, industry) and indicate the percentages of the total and federally financed portions that are considered basic research.
- **Item 1A requested** that institutions report the amount of their total and federally financed expenditures from Item 1 that were passed through to a higher education institution or other subrecipient.
- **Item 1B requested** that institutions report the amount of their total and federally financed expenditures from Item 1 that were received by the institution as a subrecipient from a higher education institution or other entity.

- **Item 2** requested total and federally financed current fund expenditures for separately budgeted R&D by detailed S&E fields. Respondents were informed that the sum of expenditures across fields should match the total and federal expenditures reported on Item 1.
- **Item 2A** requested total and federally financed current fund expenditures for separately budgeted R&D by detailed non-S&E field.
- **Item 2B** requested federally financed current fund expenditures for separately budgeted R&D by detailed S&E fields and federal agency. **Respondents** were informed that the sum of expenditures across agencies should match the federal expenditures reported in the federal column of Item 2.
- **Item 3** requested total and federally financed current fund expenditures for equipment for separately budgeted R&D by detailed S&E fields.
- **Item 4** requested the number and amount of projects awarded for R&D in both S&E and non-S&E fields by source of funding.

2.1.3. Historical Changes to the Survey Instrument

There have been several revisions to the Academic R&D Expenditures Survey throughout its 37 years of data collection. The following major changes in the survey data items occurred since 1972, the first year of data included in the accompanying data file.

- From FY 1972–77 the survey requested data on expenditures for instruction and departmental research.
- In FY 1978 the survey did not collect data on expenditures for character of work (i.e., basic and applied R&D).
- From FY 1979–89 the survey requested data on capital expenditures for facilities and equipment for research, development, and instruction. Capital expenditures may include fixed or movable furnishings and scientific equipment, facilities ancillary or integral to R&D efforts, and special separate facilities to house S&E apparatus. These expenditures may also refer to any article of nonexpendable, tangible personal property having a useful life of more than 2 years and an acquisition cost of \$500 or more per unit. This survey item was removed from the survey in 1990. Because the classification systems and the methods used to capture capital expenditures differed widely among institutions, data analysts should interpret these data with care.
- In FY 1981 a question about expenditures for research equipment by S&E field (Item 3) was added to the survey and has been included in all subsequent collections.
- Beginning in FY 1996, the survey collected data on funds passed through to subrecipients (Item 1a).
- In FY 2000 an item asking for funds received as a subrecipient was added to the survey (Item 1b).
- In FY 2003 items on non-S&E R&D expenditures (Item 2a) and R&D expenditures by sponsoring federal agency (Item 2b) were added to the survey.
- In FY 2009 an item requesting the number and amount of research awards received in both S&E and non-S&E fields by source of funding (Item 4) was added to the survey.

Some of the S&E fields in the questionnaire have also been modified:

- Beginning in FY 1980, data were collected separately for the engineering and environmental science specialties.
- In FY 1990 a metallurgical and materials engineering category was added.
- In FY 1997 a bioengineering and biomedical engineering category was added.

2.1.4. FY 2009 Pilot Survey

Forty surveyed institutions did not complete the FY 2009 Academic R&D Expenditures Survey instrument, but instead responded to a pilot version of the new HERD Survey. The HERD Survey resulted from an effort by NCSES to redesign the Academic R&D Expenditures Survey. Much of the data requested as part of the HERD Survey were the same as requested in the Academic R&D Expenditures Survey. Expenditures collected as part of the HERD Survey pilot study are included in the data files. When necessary, pilot institutions were asked to provide additional information so that a complete data set could be presented. A flag for these 40 institutions is included in the data file.

Although much of the data requested for both surveys was the same, revisions to definitions produced sizable trend changes in some cases. For example, expenditures for clinical trials and research training grants were explicitly requested on the HERD Survey but not the Academic R&D Expenditures Survey. For more information about the HERD Survey pilot study, contact the NCSES project manager.

2.1.5. Excluded Data

Some data collected as part of the R&D Expenditures Surveys are not included in the public use data files. Expenditures by fields of science are not available for FY 1972. Due to incomplete data for FY 1996 Item 1a (expenditures passed through to subrecipients), these data are not included in the public use files. Based on feedback from data collection staff and respondents, data for Item 4 (R&D projects awarded) are not included in the FY 2009 public use data file due to concerns about data quality. Information categorized as confidential is also excluded from publicly available data. Information received from individual institutions regarding the distinction between separately budgeted institutional funding and unrecovered indirect costs and cost sharing is not published or released; only aggregate totals appear in tabulations. Similarly, institution-level reports of basic, applied, and developmental research expenditures are not released to the public. Data that are collected as part of any optional survey items are not published or included in public use data files.

2.1.6. Data Element Availability by Year

The following tables list the data variables included in the Academic R&D Expenditures Survey public use files by survey year (FY 1972–2009). The FY 1978 survey collected data from doctorate-granting institutions only; NCSES estimated data for all institutions that were not sampled. These estimates are included in the first record of the FY 1978 public use file.

Data element availability is indicated by these symbols:

- A = All qualifying academic institutions
- D = Doctorate-granting institutions only
- = Not available (in most cases, not collected)

These lists should be used in conjunction with the survey questionnaire.

Item 1. Academic R&D Expenditures by Source

Line	Source of Funds	Fiscal Year		
		1972–77	1978	1979–2009
1100	Total	A	D	A
1110	Federal government	A	D	A
1125	State and local governments	A	D	A
1150	Industry	A	D	A
1160	Institution funds	A	D	A
1175	All other sources	A	D	A

Item 1A. Academic R&D Expenditures Passed Through to Subrecipients

Line	Source of Funds	Fiscal Year
		1997–2009
1900	Total	A
1910	To higher education subrecipients	A
1920	To other subrecipients	A

Item 1b. Academic R&D Expenditures Received as a Subrecipient

Line	Source of Funds	Fiscal Year
		2000–09
1600	Total	A
1610	From higher education pass-through entities	A
1620	From other pass-through entities	A

Item 2. Academic R&D Expenditures by Field

Line	Field of S&E	Fiscal Year						
		1972	1973–74	1975–77	1978	1979	1980–89	1990–2009
1400	Total	—	A	A	D	A	A	A
1410	Engineering (total)	—	A	A	D	A	A	A
1411	Aeronautical & astronautical eng.	—	—	—	—	—	A	A
1418	Bioengineering/ biomedical eng.	—	—	—	—	—	—	A
1412	Chemical eng.	—	—	—	—	—	A	A
1413	Civil eng.	—	—	—	—	—	A	A
1414	Electrical eng.	—	—	—	—	—	A	A
1415	Mechanical eng.	—	—	—	—	—	A	A
1417	Metallurgical & materials eng.	—	—	—	—	—	—	A
1416	Other	—	—	—	—	—	A	A
1420	Physical sciences (total)	—	A	A	D	A	A	A
1421	Astronomy	—	A	A	D	A	A	A
1422	Chemistry	—	A	A	D	A	A	A
1423	Physics	—	A	A	D	A	A	A
1424	Other	—	A	A	D	A	A	A
1430	Environmental sciences (total)	—	A	A	D	A	A	A
1431	Atmospheric sciences	—	—	—	—	—	A	A
1432	Earth sciences	—	—	—	—	—	A	A
1433	Oceanography	—	—	—	—	—	A	A
1434	Other	—	—	—	—	—	A	A
1441	Mathematical sciences	—	A	A	D	A	A	A
1442	Computer sciences	—	A	A	D	A	A	A
1450	Life sciences (total)	—	A	A	D	A	A	A
1451	Agricultural sciences	—	A ¹	A	D	A	A	A
1452	Biological sciences	—	A ¹	A	D	A	A	A
1453	Medical sciences	—	A	A	D	A	A	A
1454	Other	—	A	A	D	A	A	A
1460	Psychology (total)	—	A	A	D	A	A	A
1470	Social sciences (total)	—	A	A	D	A	A	A
1471	Economics	—	A	A	D	A	A	A
1472	Political science	—	A	A	D	A	A	A

Line	Field of S&E	Fiscal Year						
		1972	1973–74	1975–77	1978	1979	1980–89	1990–2009
1473	Sociology	—	A	A	D	A	A	A
1474	Other	—	A	A	D	A	A	A
1480	Other sciences, nec	—	A	A	D	A	A	A

nec = not elsewhere classified.

¹ Biological sciences data prior to FY 1974 included agricultural sciences. FY 1973 data were retroactively split into agricultural and biological sciences using the relative FY 1974 percentages reported by each institution.

Item 2A. Academic R&D Expenditures by Non-S&E Field

Line	Field of Non-S&E	Fiscal Year
		2003–09
1500	Total	A
1510	Education	A
1520	Humanities	A
1530	Law	A
1540	Visual & performing arts	A
1550	Business & management	A
1560	Communications, journalism, and library science	A
1570	Social work	A
1580	Other non-S&E, nec	A

Item 2B. Academic R&D Expenditures by Field and Agency

Line	Field of S&E	Fiscal Year
		2003–09
1400	Total	A
1410	Engineering (total)	A
1411	Aeronautical & astronautical eng.	A
1418	Bioengineering/biomedical eng.	A
1412	Chemical eng.	A
1413	Civil eng.	A
1414	Electrical eng.	A
1415	Mechanical eng.	A
1417	Metallurgical & materials eng.	A
1416	Other	A
1420	Physical sciences (total)	A
1421	Astronomy	A
1422	Chemistry	A
1423	Physics	A
1424	Other	A
1430	Environmental sciences (total)	A
1431	Atmospheric sciences	A
1432	Earth sciences	A
1433	Oceanography	A
1434	Other	A
1441	Mathematical sciences	A
1442	Computer sciences	A
1450	Life sciences (total)	A
1451	Agricultural sciences	A
1452	Biological sciences	A
1453	Medical sciences	A
1454	Other	A
1460	Psychology	A
1470	Social sciences (total)	A
1471	Economics	A
1472	Political science	A
1473	Sociology	A
1474	Other	A
1480	Other sciences, nec	A

Item 3. Current Fund Research Equipment Expenditures by Field

Line	Field of S&E	Fiscal Year		
		1981 –83	1984 –89	1990– 2009
1800	Total	A	A	A
1810	Engineering (total)	A	A	A
1811	Aeronautical & astronautical eng.	A	A	A
1818	Bioengineering/biomedical eng.	—	—	A
1812	Chemical eng.	A	A	A
1813	Civil eng.	A	A	A
1814	Electrical eng.	A	A	A
1815	Mechanical eng.	A	A	A
1817	Metallurgical & materials eng.	—	—	A
1816	Other	A	A	A
1820	Physical sciences (total)	A	A	A
1821	Astronomy	A	A	A
1822	Chemistry	A	A	A
1823	Physics	A	A	A
1824	Other	A	A	A
1830	Environmental sciences (total)	A	A	A
1831	Atmospheric sciences	A	A	A
1832	Earth sciences	A	A	A
1833	Oceanography	A	A	A
1834	Other	A	A	A
1841	Mathematical sciences	A	A	A
1842	Computer sciences	A	A	A
1850	Life sciences (total)	A	A	A
1851	Agricultural sciences	A	A	A
1852	Biological sciences	A	A	A
1853	Medical sciences	A	A	A
1854	Other	A	A	A
1860	Psychology (total)	A	A	A
1870	Social sciences (total)	A	A	A
1871	Economics	A	A	A
1872	Political science	A	A	A
1873	Sociology	A	A	A
1874	Other	A	A	A
1880	Other sciences, nec	A	A	A

Capital Expenditures by Area

Line	Field of S&E	Fiscal Year
		1979-89
1700	Total	A
1710	Engineering	A
1720	Physical sciences	A
1730	Environmental sciences	A
1740	Mathematical & computer sciences	A
1750	Life sciences	A
1760	Psychology	A
1770	Social sciences	A
1780	Other sciences, nec	A

2.2. DATA IMPUTATION

Imputation for this version of the survey involved determining inflator and deflator factors from fully responding institutions for four key variables:

- Total R&D expenditures
- Federally financed R&D expenditures
- Total research equipment expenditures
- Total capital expenditures, where applicable

For nonrespondent institutions, these variables were estimated by applying the appropriate inflator and deflator to the previous year's data values for that institution. Inflators and deflators were based on the institution's peer group with respect to highest degree granted and type of control (i.e., public or private).

The key variables were then distributed among the various subtotal and detailed fields using the relative percentages that were last reported by that institution. If no percentages were available, the summary percentages for the institution's peer group were used.

2.2.1. Retroimputation of Prior Years' Data

A number of institutions in the survey universe are intermittent respondents; they provide data one year, do not respond in one or more subsequent years, and then provide data again. Data for the years in which no response was received were imputed as described in the section above. While the imputation algorithm accurately reflects national trends, it cannot account for reporting anomalies at individual institutions. For this reason, a separate retroimputation for FY 1972–2008 data was performed following the FY 2009 imputation.

For each intermittently reporting institution, data were retroimputed whenever an institution that had not responded in the previous year(s) responded in FY 2009. For example, if data were reported for FY 2005 and FY 2009 but not for the intervening years, the difference between the reported figures for each item total was calculated, and these amounts were then evenly distributed across the intervening years. The new figures were spread across disciplines or sources of support on the basis of the most recent reporting pattern. These procedures resulted in much more consistent reporting trends for individual institutions but had little effect on aggregated figures reflecting national totals.

For more detailed information on imputation and retroimputation procedures for the Academic R&D Expenditures Survey, see Section 4 of the FY 2009 Survey Methodology Report.

2.3. PUBLIC USE DATA FILES

After completion of the annual data processing cycle, the collected data are stored in files for public use and archival purposes. These files include any corrections to prior-year data that survey respondents have submitted. For each survey year, the data file contains two record types: institution header records and data records. Apart from alphabetic information (e.g., names of

institutions, cities), most identifying information in the survey data files is represented by numeric codes. The allowed values, their meanings, and file formats are described below.

2.3.1. Institution Header Records

The institution header record provides basic identifying data for a given institution. One record of this type exists for each institution for each survey cycle in which the institution participated; this record is always the first record for an institution in a given year. The format is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–12	Combined ID (“000000” if not to be combined)
13–16	Fiscal Year
17	Record Type 1 = Institution header record
18	HBCU Indicator F = not HBCU T = HBCU Blank = No indicator assigned; aggregation of institutions
19	Medical School Indicator F = Does not have a medical school N = Null; information was not included T = Has a medical school
20	High Hispanic Enrollment Indicator F = Is not a high Hispanic enrollment institution N = Null; information was not included T = Is a high Hispanic enrollment institution
21	Type of Institution 1 = Academic
22	Highest S&E Degree Granted by Institution 1 = Doctorate 2 = Master’s 3 = Bachelor’s 4 = No S&E Degree (may grant bachelor’s or higher in nonscience program) 8 = 2-Year Program 9 = No degree assigned; aggregation of institutions

23	Type of Control 1 = Public 2 = Private ? = No institutional control assigned; aggregation of institutions
24	Included in FY 2009 Pilot Survey T = True F = False
25–74	Institution Name
75–94	City Name
95–96	State Abbreviation ?? = No state; aggregation of institutions
97–106	ZIP Code ???? = No ZIP Code; aggregation of institutions

2.3.2. Data Records

Data records contain the numerical responses to the Academic R&D Expenditures Survey questionnaire items. These records are identified by line numbers corresponding to the line numbers on the FY 2009 questionnaire. Although the meaning of some line numbers changed during the period from FY 1972 to FY 2009, the line references for these years have been made consistent with the FY 2009 format to facilitate data trend analyses. All data items are expressed in thousands of dollars. A data record is present for each data line on the questionnaire to which a nonzero response has been received. For total lines (i.e., line numbers that end in “00”) in each questionnaire item, a data record is present regardless of whether a zero or nonzero response has been received. Since each questionnaire line can have up to nine columns of data, the data record has nine columns of responses. A status code associated with each data column indicates the source of or special treatment for the data. The format is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–12	Combined Code
13–16	Fiscal Year
17	Record Type 2 = Data record

18–21	Questionnaire Number
22–76	Question Content Description
77–131	Questionnaire Row Description
132–161	Question Column Description
162–169	Data Expenditure amount
170	Status Code for Data Blank = Normal Response e = Estimated by NCSES i = Imputed by Computer for Nonresponse n = Data Not Available

3. ADDITIONAL RESOURCES

3.1. Publications

The data from this survey are summarized in an *InfoBrief* and published annually in detailed statistical tables in the series Higher Education Research and Development, available by fiscal year on the NCSES website. Data for major data elements are available starting in 1972. Information from the survey is also included in Science and Engineering Indicators, National Patterns of R&D Resources, Science and Engineering State Profiles, and Academic Institutional Profiles.¹

3.2. Electronic Access

All data from this survey are available on the NCSES website. Selected aggregate data are provided in public use data files upon request.

3.3. Contact for More Information

Additional information about this survey can be obtained by contacting:

Michael Gibbons

Research and Development Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
2415 Eisenhower Avenue
Alexandria, VA 22314

Phone: (703) 292-4590

E-mail: mgibbons@nsf.gov

¹ Publications from NCSES can be found at <http://www.nsf.gov/statistics/>

4. OTHER FORMATS

4.1. HERD (FY 2010 – Present)

4.1.1. Excel Data File Format

The institutional characteristics provide basic identifying information for a given institution. Data records contain the numerical responses to the HERD Survey questionnaire items. These records are identified by survey question and row numbers corresponding to those on the FY 2018 questionnaire. It should be noted that although the meaning of some question and row numbers changed during the period from FY 2010 to FY 2018, the references for these years have been made consistent with the FY 2018 format to facilitate data trend analyses. All expenditure data items are expressed in thousands of dollars. Data for survey question 15 and 16 are headcounts and are expressed as actual numbers. For each data line on the questionnaire to which a nonzero response has been received, a data record is present. For total rows in each survey question, a data record is present regardless of whether a zero or nonzero response has been received. A status code associated with each data column indicates the source of and/or special treatment for the data.

For FY 2012 to FY 2018 data from institutions that completed the short form version of the survey are published separately from those that completed the standard survey form, and for this reason the two populations are treated separately in these data files as well.

The format for the data reported by institutions responding to the standard survey form is as follows:

COLUMN	CONTENTS
inst_id	ID Code
year	Fiscal Year
nces_inst_id	NCSES Institution ID
ipeds_unitid	IPEDS Institution ID
hbcu_flag	HBCU Indicator 0 = not HBCU 1 = HBCU Blank = No indicator assigned; aggregation of institutions
med_sch_flag	Medical School Indicator F = Does not have a medical school N = Null; information was not included T = Has a medical school

hhe_flag	High Hispanic Enrollment Indicator 0 = Is not a high Hispanic enrollment institution 1 = Is a high Hispanic enrollment institution Blank = No indicator assigned; aggregation of institutions
toi_code	Type of Institution 1 = Academic
hdg_code	Highest Degree Granted by Institution 1 = Doctorate 2 = Master's 3 = Bachelor's 4 = Associate's 5 = No degree 6 = Professional degree
toc_code	Type of Control 1 = Public 2 = Private ? = No institutional control assigned; aggregation of institutions
inst_name_long	Institution Name
inst_city	City Name
inst_state_code	State Abbreviation ?? = No state; aggregation of institutions
inst_zip	ZIP Code ???? = No ZIP Code; aggregation of institutions
questionnaire_no	Survey Questionnaire Number
question	Survey Question Content Description
row	Survey Question Row Description
column	Survey Column Description
data	Data Question 1.1 Response Codes -1=Don't Know 0=No 1=Yes

status	Status Code for Data Blank = Normal Response i = Imputed by Computer for Nonresponse n = Data Not Available
othinfo	Other Information Reason Specific Types of Funding listed in Question 1.1 were not Included in Data Reported for Current Year (FY 2012 and later years) Specific Federal Agencies Listed in Survey Question 10 Inclusion of Clinical Trials in FY 2009 Submission for Question 5.1 (FY 2010 files only)
othinfo_s	Status Code for Data Blank = Normal Response i = Imputed by Computer for Nonresponse n = Data Not Available
standardized_agency_name	Standardized Name for Specific Federal Agency Listed in Question 10

The format for the data reported by institutions responding to the short survey form is as follows:

COLUMN	CONTENTS
inst_id	ID Code
year	Fiscal Year
nces_inst_id	NCSES Institution ID
ipeds_unitid	IPEDS Institution ID
hbcu_flag	HBCU Indicator 0 = not HBCU 1 = HBCU Blank = No indicator assigned; aggregation of institutions
med_sch_flag	Medical School Indicator F = Does not have a medical school N = Null; information was not included T = Has a medical school
hhe_flag	High Hispanic Enrollment Indicator

0 = Is not a high Hispanic enrollment institution
 1 = Is a high Hispanic enrollment institution
 Blank = No indicator assigned; aggregation of institutions

toi_code	Type of Institution 1 = Academic
hdg_code	Highest Degree Granted by Institution 1 = Doctorate 2 = Master's 3 = Bachelor's 4 = Associate's 5 = No degree 6 = Professional degree
toc_code	Type of Control 1 = Public 2 = Private ? = No institutional control assigned; aggregation of institutions
inst_name_long	Institution Name
inst_city	City Name
inst_state_code	State Abbreviation ?? = No state; aggregation of institutions
inst_zip	ZIP Code ???? = No ZIP Code; aggregation of institutions
questionnaire_no	Survey Questionnaire Number
question	Survey Question Content Description
row	Survey Question Row Description
column	Survey Question Column Description
data	Data Question 1.1 Response Codes -1=Don't Know 0=No 1=Yes
status	Status Code for Data Blank = Normal Response

i = Imputed by Computer for Nonresponse
n = Data Not Available

othinfo

Other Information

Reason Specific Types of Funding listed in Question 1.1 were not
Included in Data Reported for Current Year (FY 2012 and later years)

4.1.2. SAS Data File Format

SAS format follows the Excel format. See section 4.1.1.

4.1.3. SPSS Data File Format

SPSS format follows the Excel format. See sections 4.1.1.

4.1.4. CSV Data File Format

CSV format follows the Excel format. See sections 4.1.1.

4.2 Academic R&D Survey (FY 1979 – FY 2009)

4.2.1 Excel Data File Format

The institutional characteristics provide basic identifying information for a given institution. Data records contain the numerical responses to the survey questionnaire items. These records are identified by line numbers corresponding to the line numbers on the FY 2009 questionnaire. It should be noted that although the meaning of some line numbers has changed during the period from FY 1972 to FY 2009, the line references for these years have been made consistent with the FY 2009 format to facilitate data trend analyses. All data items are expressed in thousands of dollars. For each data line on the questionnaire to which a nonzero response has been received, a data record is present. For total lines (i.e., line numbers that end in “00”) in each questionnaire item, a data record is present regardless of whether a zero or nonzero response has been received. Since each questionnaire line can have up to nine columns of data, the data record has nine columns of responses. A status code associated with each data column indicates the source of and/or special treatment for the data. The data file format is as follows:

COLUMN	CONTENTS
ficie	ID Code
ficie_combined	Combined ID (“000000” if not to be combined)
year	Fiscal Year
hbcu_flag	HBCU Indicator F = not HBCU T = HBCU

Blank = No indicator assigned; aggregation of institutions

has_med_schl_flag	Medical School Indicator F = Does not have a medical school N = Null; information was not included T = Has a medical school
hhe_flag	High Hispanic Enrollment Indicator F = Is not a high Hispanic enrollment institution N = Null; information was not included T = Is a high Hispanic enrollment institution
toi_code	Type of Institution 1 = Academic
hdg_code	Highest S&E Degree Granted by Institution 1 = Doctorate 2 = Master's 3 = Bachelor's 4 = No S&E Degree (may grant bachelor's or higher in nonscience program) 8 = 2-Year Program 9 = No degree assigned; aggregation of institutions
toc_code	Type of Control 1 = Public 2 = Private ? = No institutional control assigned; aggregation of institutions
pilot_fy09_flag	Included in FY 2009 Pilot Survey T = True F = False
inst_name_long	Institution Name
inst_city	City Name
inst_state	State Abbreviation ?? = No state; aggregation of institutions
inst_zip	ZIP Code

???? = No ZIP Code; aggregation of institutions

questionnaire_no	Survey Questionnaire Number
question	Survey Question Content Description
row	Survey Question Line Description
column	Survey Question Column Description
data	Data
status	Status Code for Data Column 1 Blank = Normal Response e = Estimated by NCSES i = Imputed by Computer for Nonresponse n = Data Not Available

4.2.2. SAS Data File Format

SAS format follows the Excel format. See sections 4.2.1.

4.2.3. SPSS Data File Format

SPSS format follows the Excel format. See sections 4.2.1.

4.2.4. CSV Data File Format

CSV format follows the Excel format. See sections 4.2.1.