Guide for Public Use Data Files

National Science Foundation's FFRDC Research and Development Survey:

Fiscal Year 2013

September 2014

National Science Foundation
National Center for Science and Engineering Statistics
Research and Development Statistics Program

Table of Contents

1. SURVEY POPULATION AND FRAME 3

2. FFRDC R&D SURVEY (FY2010 – PRESENT) 4

2.1. Survey Instrument and Data Availability 4
2.1.1. Key Variables 4
2.1.2. Changes to the Survey for FY 2013 5
2.1.3. Excluded Data 5
2.2. Public Use Data Files 5
2.2.1. Institution Header Records 5
2.2.2. Data Records 6

3. FFRDC R&D EXPENDITURES SURVEY (FY1979 – FY2009) 7

3.1. Survey Instrument and Data Availability 7
3.1.1. Key Variablest 7
3.1.2. Historical Changes to the Survey Instrument 7
3.1.3. Excluded Data 8
3.1.4. Data Element Availability by Year 8
3.2. Public Use Data Files 13
3.2.1. Institution Header Records 13
3.2.2. Data Records 13

4. ADDITIONAL RESOURCES 14

4.1. Publications 14
4.2. Electronic Access 15
4.3. Contact for More Information 15

5. OTHER FORMATS 15

5.1. FFRDC R&D SURVEY (FY2010 – Present) 15
5.1.1. Excel Data File Format 15
5.1.2. SAS Data File Format 17

5.1.3. SPSS Data File Format 17
5.1.4. CSV Data File Format 17
5.2. FFRDC R&D EXPENDITURES SURVEY (FY1979 – FY2009) 17
5.2.1. Excel Data File Format 17
5.2.2. SAS Data File Format 18
5.2.3. SPSS Data File Format 18
5.2.4. CSV Data File Format 18

This data users' guide provides a detailed description of the public use data files for the FY 2013 National Science Foundation (NSF) FFRDC Research and Development Survey (FFRDC R&D Survey). The FFRDC R&D Survey replaces the previous annual collection, the Survey of Research and Development Expenditures at Federally Funded Research and Development Centers (FFRDC R&D Expenditures Survey), which was conducted from FY 1953 through FY 2009. The second and third section of this guide provides a detailed description of the public use data files for each version of the R&D survey. Each of these sections includes an overview of the two surveys and a detailed description of each data file structure. For additional information about the survey population and procedures refer to the FY 2013 Survey Methodology Report and methodology reports from previous collection cycles (available upon request from the NSF project manager listed below (4.3).

1. SURVEY POPULATION AND FRAME

The FFRDC R&D Survey has been an annual census of the population of eligible FFRDCs since FY 2001. Prior to FY 2001, only FFRDCs administered by academic institutions were included in this survey. FFRDCs are identified through the NSF master list of FFRDCs. NSF is responsible for maintaining the master list and queries all federal agencies annually to determine changes, additions, or deletions to the list.

The number of FFRDCs included annually in the survey population since 1979 is as follows:

Fiscal Year	Number of FFRDCs
1979	19
1980	19
1981	19
1982	19
1983	19
1984	17
1985	17
1986	17
1987	18
1988	18
1989	18
1990	18
1991	19
1992	19
1993	19
1994	18
1995	18
1996	18
1997	18
1998	17
1999	17
2000	16
2001[+]	36

Fiscal Year	Number of FFRDCs
2002+	36
2003+	36
2004+	36
2005+	37
2006+	37
2007+	38
2008+	38
2009+	39
2010+	39
2011+	40
2012+	39
2013+	40

+ This year included FFRDCs administered by industrial organizations and FFRDCs administered by nonprofit organizations.

For a complete list of FY 2013 surveyed institutions, see Appendix A of the FY 2013 Survey Methodology Report.

2. FFRDC R&D SURVEY (FY2010 – Present)

2.1. Survey Instrument and Data Availability

With the introduction of the FFRDC R&D Survey in FY 2010, a number of changes were made to the questionnaire. Three new questions were added to the survey and clinical trials and research training grants were explicitly included in the definition of R&D.

2.1.1. Key Variables

The FY 2013 FFRDC R&D Expenditures Survey consisted of five questions requesting information on expenditures from current funds for separately budgeted R&D activities during the organization's 2013 fiscal year.

- **Question 1** requested that organizations report their total current expenditures for separately budgeted R&D by source of funds.
- **Question 2** requested that institutions report the amount of their federal R&D expenditures from Question 1 that was funded by the American Recovery and Reinvestment Act of 2009 (ARRA).
- **Question3** requested the amount of federal and nonfederal R&D expenditures that were considered basic research, applied research, and/or development.
- **Question 4** requested that institutions report their total R&D expenditures by type of direct and indirect cost.
- **Question 5** requested the institutions total operating budget, excluding capital construction costs

In addition, **Question 6** 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.

2.1.2. Changes to the Survey for FY 2013

There were no changes to the data collection instrument for FY 2013.

2.1.3. Excluded Data

Some data collected as part of the FFRDC R&D Surveys are not included in the public use data files. Information categorized as confidential is excluded from publically available data. Information received from individual institutions regarding cost elements of R&D, their total operating budget, and institution contact information is not published or released.

2.2. 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, together with their meanings, and file formats are described below.

2.2.1. 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
7–10	Fiscal Year
11	Record Type = 1
12	Type of Institution 3 = FFRDC

13	FFRDC Type 0 = Not an FFRDC 1 = Administered by Academic Institution 2 = Administered by Industrial Organization 3 = Administered by Nonprofit Organization
14–88	Institution Name
89–108	City Name
109–110	State Abbreviation
111–115	ZIP Code

2.2.2. Data Records

Data records contain the numerical responses to the survey questionnaire items. These records are identified by survey question and row numbers corresponding to the numbers on the FY 2013 questionnaire. 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 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. The format is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–10	Fiscal Year
11	Record Type = 2
12–14	Survey Question Number
15–17	Survey Question Row Number
18–25	Data Column 1 Total Expenditures for that Survey Question and Row Number
26	Status Code for Data Column 1 Blank = Normal Response e = Estimated by NSF i = Imputed by Computer for Nonresponse n = Data Not Available
27–34	Data Column 2 Federal Expenditures for that Survey Question and Row Number

35	Status Code for Data Column 2
36–43	Data Column 3 Nonfederal Expenditures for that Survey Question and Row Number
44	Status Code for Data Column 3

3. FFRDC R&D EXPENDITURES SURVEY (FY1979 – FY2009)

3.1. Survey Instrument and Data Availability

3.1.1. Key Variables

The FY 2009 FFRDC R&D Expenditures Survey consisted of two items requesting information on R&D expenditures from current funds for separately budgeted R&D activities during the institutions' 2009 fiscal year:

- **Item 1** requested that institutions report their total current expenditures for separately budgeted S&E R&D expenditures by source of funds.
- **Item 2** requested the amount of total expenditures and federally financed expenditures that were considered basic research, applied research, and/or development.

3.1.2. Historical Changes to the Survey Instrument

Throughout the 36 years of data collection there were several revisions to the survey. The following major changes in the survey data items occurred since 1979, the first year of data included in the accompanying data file.

- 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. It was removed from the FFRDC survey in 2002.
- Beginning in FY 1996, the survey collected data on funds passed through to subrecipients (Item 1a). This item was removed from the FFRDC survey in 2002.
- In FY 2000 an item asking for funds received as a subrecipient was added to the survey (Item 1b). This item was removed from the FFRDC survey in 2002.
- In FY 2007 the character of work question was expanded on the FFRDC version of the survey to include basic research, applied research, and development.

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/biomedical engineering category was added.

3.1.3. Excluded Data

Some data collected as part of the R&D Expenditures Surveys are not included in the public use data file. Institution level data collected prior to FY 1979 have not been integrated with current databases and cannot be included in the public use files. Due to incomplete data for FY 1996 Item 1a, expenditures passed through to subrecipients, these data are not included in the public use files. Information categorized as confidential is also excluded from publically 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 for FY 2009 and earlier years are not released to the public. Also, data that are collected as part of an optional survey are not published or included in public use data files.

3.1.4. Data Element Availability by Year

The following tables lists the data variables include in the public use files by survey year and population type. Data cells that were collected but do not correspond to the current files are not included. Availability is indicated by these symbols:

A = All qualifying FFRDCs
— = Not available (in most cases, not collected)

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

Item 1. R&D Expenditures by Source

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

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

Line	Source of Funds	Fiscal Year	
		1997–2001	2002–09
1900	**Total**	A	—
1910	To higher education subrecipients	A	—
1920	To other subrecipients	A	—

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

Line	Source of Funds	Fiscal Year	
		2000–01	2002–09
1600	**Total**	A	—
1610	From higher education pass-through entities	A	—
1620	From other pass-through entities	A	—

Item 2. R&D Expenditures by Field

Line	Field of S&E	Fiscal Year			
		1979	1980–89	1990–2001	2002–09
1400	**Total**	A	A	A	—
1410	Engineering	A	A	A	—
1411	Aeronautical and astronautical	—	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 and materials eng.	—	—	A	—
1416	Other	—	A	A	—
1420	Physical sciences	A	A	A	—
1421	Astronomy	A	A	A	—
1422	Chemistry	A	A	A	—
1423	Physics	A	A	A	—
1424	Other	A	A	A	—
1430	Environmental sciences	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	A	—
1442	Computer sciences	A	A	A	—
1450	Life sciences	A	A	A	—
1451	Agricultural sciences	A	A	A	—
1452	Biological sciences	A	A	A	—
1453	Medical sciences	A	A	A	—
1454	Other	A	A	A	—
1460	Psychology	A	A	A	—
1470	Social sciences	A	A	A	—
1471	Economics	A	A	A	—

Line	Field of S&E	Fiscal Year			
		1979	1980–89	1990–2001	2002–09
1472	Political science	A	A	A	—
1473	Sociology	A	A	A	—
1474	Other	A	A	A	—
1480	Other sciences, nec	A	A	A	—

nec = not elsewhere classified.

Item 3. Current Fund Research Equipment Expenditures by Field

Line	Field of S&E	Fiscal Year			
		1981–83	1984–89	1990–2001	2002–09
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	—

Line	Field of S&E	Fiscal Year			
		1981–83	1984–89	1990–2001	2002–09
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 (total)	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

3.2. 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, together with their meanings, and file formats are described below.

3.2.1. 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 = 2
18	Type of Institution 3 = FFRDC
19	FFRDC Type 0 = Not an FFRDC 1 = Administered by Academic Institution 2 = Administered by Industrial Organization 3 = Administered by Nonprofit Organization
20–94	Institution Name
95–114	City Name
115–116	State Abbreviation
117–126	ZIP Code

3.2.2. Data Records

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 format is as follows:

COLUMN	CONTENTS
1–6	ID Code
7–12	Combined Code
13–16	Fiscal Year
17	Record Type = 2
18–21	Questionnaire Line Number
22–29	Data Column 1 Total Expenditures for that Line Number
30	Status Code for Data Column 1 Blank = Normal Response e = Estimated by NSF i = Imputed by Computer for Nonresponse n = Data Not Available
31–38	Data Column 2 Federal Expenditures for that Line Number
39	Status Code for Data Column 2
40–47	Data Column 3 All Other Capital Expenditures for Line Numbers 1700-1780
48	Status Code for Data Column 3

4. ADDITIONAL RESOURCES

4.1. Publications

The data from this survey are summarized in an *InfoBrief* and are published annually in the

detailed statistical tables. Data for FY 2008–13 of the FFRDC R&D Survey are available at http://www.nsf.gov/statistics/ffrdc/. Data from FY 2008 and prior years of the FFRDC R&D Expenditures Survey are included in the Academic R&D Expenditures series of detailed statistical tables, available by fiscal year at http://www.nsf.gov/statistics/rdexpenditures/. Data for major elements are available starting in 1972.

Information from the survey is also included in the series *Science and Engineering Indicators* and *National Patterns of R&D.*

4.2. 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. Special-purpose data files with selected data variables can be produced to meet particular analytical needs. Individuals interested in special-purpose files should contact the NSF project manager.

4.3. Contact for More Information

Additional information about this survey can be obtained by contacting:

Ronda Britt
Research and Development Statistics Program
National Center for Science and Engineering Statistics
National Science Foundation
4201 Wilson Boulevard, Suite 965
Arlington, VA 22230

Phone: (703) 292-7765
E-mail: rbritt@nsf.gov

5. Other Formats

5.1. FFRDC R&D SURVEY (FY2010 – Present)

5.1.1. Excel Data File Format

The institutional characteristics provide basic identifying information for a given institution. Data records contain the numerical responses to the survey questionnaire items. These records are identified by survey question and row numbers corresponding to the numbers on the FY 2013 questionnaire. 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 in survey question, a data record is present regardless of whether a zero or nonzero response has been received. Since each survey question can have up to three columns of data, the data record has three 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
inst_id	ID Code
year	Fiscal Year
toi_code	Type of Institution 3 = FFRDC
ffrdctype	FFRDC Type 0 = Not an FFRDC 1 = Administered by Academic Institution 2 = Administered by Industrial Organization 3 = Administered by Nonprofit Organization
inst_name_long	Institution Name
inst_city	City Name
inst_state	State Abbreviation
inst_zip	ZIP Code
question	Survey Question Number
row	Survey Question Row Number
data1	Data Column 1 Total Expenditures for that Survey Question and Row Number
status1	Status Code for Data Column 1 Blank = Normal Response e = Estimated by NSF i = Imputed by Computer for Nonresponse n = Data Not Available
data2	Data Column 2 Federal Expenditures for that Survey Question and Row Number
status2	Status Code for Data Column 2
data3	Data Column 3 Nonfederal Expenditures for that Survey Question and Row Number
status3	Status Code for Data Column 3

5.1.2. SAS Data File Format

SAS format follows the Excel format. See sections 5.1.1.

5.1.3. SPSS Data File Format

SPSS format follows the Excel format. See sections 5.1.1.

5.1.4. CSV Data File Format

CSV format follows the Excel format. See sections 5.1.1.

5.2. FFRDC R&D EXPENDITURES SURVEY (FY1979 – FY2009)

5.2.1. 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 had 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
fice	ID Code
fice_combined	Combined ID ("000000" if not to be combined)
year	Fiscal Year
toi_code	Type of Institution 3 = FFRDC

ffrdctype	Type of FFRDC 0 = Not an FFRDC 1 = Administered by Academic Institution 2 = Administered by Industrial Organization 3 = Administered by Nonprofit Organization
inst_name_long	Institution Name
inst_city	City Name
inst_state	State Abbreviation
inst_zip	ZIP Code
line	Questionnaire Line Number
data1	Data Column 1 Total Expenditures for that Line Number
status1	Status Code for Data Column 1 Blank = Normal Response e = Estimated by NSF i = Imputed by Computer for Nonresponse n = Data Not Available
data2	Data Column 2 Federal Expenditures for that Line Number
status2	Status Code for Data Column 2
data3	Data Column 3 All Other Capital Expenditures for Line Numbers 1700-1780
status3	Status Code for Data Column 3

5.2.2. SAS Data File Format

SAS format follows the Excel format. See sections 5.2.1.

5.2.3. SPSS Data File Format

SPSS format follows the Excel format. See sections 5.2.1.

5.2.4. CSV Data File Format

CSV format follows the Excel format. See sections 5.2.1.