Information on the Data Archive for the STEM Workforce Data Project
(Updated 10/04/07)

Data Archives 1 to 4 (Group 1 Tables) provide information on employed persons by detailed occupation and sex, for 1983 to 2002. The data were supplied by staff at the U.S. Bureau of Labor Statistics; our thanks and appreciation to those who assembled these data and who helped to direct us to it. Data Archive 1 contains annual average employment estimates by detailed occupational categories for 1983 through 2002 for the entire U.S. labor force (aged 16 or older). Some occupational categories may have been combined with others in order to yield consistent groups across the entire time series. Data Archive 2 is similar but reports only on men; Data Archive 3 is similar but reports only on women; Data Archive 4 provides the percentage women results for these data.

Comments on the original BLS data: This is the most consistently defined and maintained statistical time series that we have been able to locate for the employment of technical professionals in the United States during the final years of the 20th Century.

Data Archives 5 to 7 (Group 2 Tables) provide data on employed persons by detailed scientific, technological, engineering and mathematical (STEM) occupations and sex, for 1983 to 2002. These tables reflect a reorganization of the original BLS data to yield information on STEM occupations. Data archive 5 provides data for the entire workforce; data archive 6 provides data for women in STEM occupations; and data archive 7 shows women as a percent of STEM workforce.

Comments on the formation of STEM occupational groups: "Postsecondary teachers" -- college faculty -- are moved to appropriate disciplinary groups. Some STEM occupations consist almost entirely of faculty, such as historians. Historians also illustrate judgments we have made about whether or not fields should be treated as sciences. In general -- not always -- we have followed federal occupational classification conventions, which group history with other social sciences. One obvious exception to this practice comes with programmers, which had been treated as technicians in the past, but which (as with our similar work on the Information Technology or "IT" workforce) we think should be grouped with other computer professionals -- a judgment also incorporated in the newest versions of the Standard Occupational Codes. Another exception is our inclusion of sales engineers, picked up from the SOC sales occupations group, with "Engineers n.e.c.," on the grounds that sales engineers are often highly trained technical professionals whose skills help customers select optimal solutions to their needs.

For data archives 6 and 7, figures for a number of the smaller STEM professions are not sufficient to support calculations on percentages for women's share of employment, including physicists and astronomers, actuaries, statisticians, mathematical scientists n.e.c, metallurgical and materials engineers, mining engineers, petroleum engineers, nuclear engineers, agricultural engineers, marine engineers and naval architects, postsecondary engineering teachers, surveyors and mapping scientists, industrial and mechanical engineering technicians, and sociologists. Data on women's share of jobs in a number of other professions are available only for some, but not all, years in the time series.

We should also note what is *not* in these data. An SOC occupational category for "science, engineering, and computer systems managers" has been defined for some time but was not usually separated out from other kinds of specialty managers in CPS results during the 1980's and 1990's. Architects have been subtracted out of the "engineers, surveyors, and architects" aggregate.

In these tables, unique estimates are generated by BLS for both specific occupational categories and for aggregated groups. Due to rounding to the nearest thousand, aggregate estimates will not

necessarily match the exact total for their components. Indeed, because the aggregates reflect larger sample sizes, those totals are the most reliable and have been preserved wherever feasible.

If no estimate has been provided by BLS, such cases are flagged with a dash (" -- "); if estimates are available but the number is less than 500, the results, which are stated in thousands of persons, will be rounded down to zero.

Data archives 8 to 17 (Group 3 Tables) present data on employed and experienced unemployed persons by detailed occupation, sex, race, and Hispanic origin, annual average for 1994 to 2003, based on the Bureau of Labor Statistics' Current Population Survey (CPS). Each file presents annual averages for one year. These tables are also available (in the same format) from the Bureau of Labor Statistics at ftp://ftp.bls.gov/pub/special.requests/lf/. Data from these archives were used to prepare the second STEM Workforce report, ***Women in Science and Technology***.

Data archive 18 (Group 4 Table) provides data on employed and unemployed persons by detailed STEM occupation, sex and race/ethnicity for 2003. STEM occupational groups were formed in the same manner described for data archives 5 to 7. Data archive 18 also contains comments on changes in SOC codes for 2003. This file was used to prepare the first STEM Workforce report, ***Twenty Years of Scientific and Technical Employment***.

Data archives 19 to 27, 53 to 54, and 105 (Group 5 Tables) present data on median weekly earnings by occupation and sex for 1995 to 2006, based on the Bureau of Labor Statistics' Current Population Survey (CPS). Each file presents annual averages for one year. These tables are also available (in the same format) from the Bureau of Labor Statistics at ftp://ftp.bls.gov/pub/special.requests/lf/. Data from these archives were used to prepare the second STEM Workforce report, ***Women in Science and Technology***, and the fifth report, ***Science and Technology Salaries: Trends and Details, 1995-2005***.

Data archives 28 to 40 (Group 6 Tables) present data on minorities employed in STEM occupations from 1994 to 2004. Data archive 28 provides summary statistics for the entire period, along with a detailed explanation of how the statistics were assembled. The data for these archives are based on the Bureau of Labor Statistics' Current Population Survey (CPS). Data from these archives were used to prepare the third STEM Workforce report, ***Sisyphus Revisited: Participation by Minorities in STEM Occupations, 1994-2004***.

Data archives 41 to 46 (Group 7 Tables) present data on foreign born workers in STEM occupations in the U.S. from 1994 to 2002. Data from these archives were used to prepare the fourth STEM Workforce report, ***The Foreign Born in Science and Technology***. Data archives 41 to 44 are also included in the printed report, while data archives 45 and 46 contain actual numbers of workers and foreign born workers in detailed STEM occupations for 1994 to 2002. This set of data archives is based on data from the merged outgoing rotation group (MORG) files generated by the National Bureau of Economic Research. This public use microdata combines independent random samples based on the Bureau of Labor Statistics' Current Population Survey (CPS).

Data Archives 47 to 54 (Group 8 Tables) include the source data for Report Number 5, ***Science and Technology Salaries: Trends and Details, 1995-2005***. Data archive 47 contains the original source data. Data archive 48 includes the numbers of cases, in thousands, for the medians in data archive 47. Data archive 49 is the conversion of median weekly earnings for full time workers to annual salaries (sheet 1 data X 52 weeks). Data archive 50 contains the conversion of median annual salaries for full time workers to 2005 dollars (CPI adjustment). Data archive 51 includes the conversion of median annual salaries for full time workers to 2005 dollars (CPI adjustment), rounded to the nearest $500. Data archive 52 is the check on rounding. Data archives 47 to 52 make use of a set of special tables, prepared by the U.S. Bureau of Labor Statistics, which merge results from monthly Current Population Surveys

(CPS) into sets of annual estimates. One of these, the 'aat39' series, is a source of data on earnings. The statistics were originally acquired from the BLS' FTP website, ftp://ftp.bls.gov/pub/special.requests/lf/, and are also included in our data archive in the Group 5 Tables, as **data archives 19 to 27, 53 and 54**.

Data Archives 55 to 66 (Group 9 Tables) present data on degrees awarded by level, broad field and sex, from 1966 to 2004. These data were derived from the National Science Foundation's WebCASPAR Database, and provide supporting material for Report Number 6, ***Four Decades of STEM Degrees, 1966-2004: "The Devil is in the Details."***

Data Archives 67 to 78 (Group 10 Tables, bachelor's degrees) and **Data Archives 79 to 90** (Group 11 Tables, master's degrees) show degrees awarded by broad field, race/ethnicity and citizenship, for 1977 to 2004. These data were also derived from the National Science Foundation's WebCASPAR Database. Data by race/ethnicity and citizenship were not available prior to 1977, and only in selected years until 1989. No data were available for 1999.

Data Archives 91 to 101 (Group 12 Tables), also from NSF's WebCASPAR Database, show doctorates awarded by broad field, race/ethnicity and citizenship for 1973 to present. Data by race/ethnicity and citizenship at the doctorate level were not available prior to 1973.

Data Archives 102 to 104 (Group 13 Tables) provide supporting material for Report Number 7: ***STEM Employment Forecasts and Distribution Among Employment Sectors***. Data Archive 102 shows employment in STEM by occupation and industry in 2004, projected employment in STEM by occupation and industry in 2014 and the projected percent change. Data Archive 103 presents employment in STEM by occupation in 2004 and projected employment in STEM by occupation in 2014. This archive is ranked by the projected percent change. Data Archive 104 presents the same information ranked by the projected number change. All three of these data archives were derived from the Bureau of Labor Statistics National Employment Matrix.

Data Archives 106 and 107 (Group 14 Tables) provide supporting material for Report Number 6: ***Is U.S. Science and Technology Adrift?*** Data Archive 106 provides employment data for 1983 to 2006 and Data Archive 107 examines salary data for 1995 to 2006.

For More Information:

Lisa Frehill, Executive Director
Commission on Professionals in Science and Technology
1200 New York Avenue, NW
Suite 113
Washington, DC 20005
(202) 326-7084
lfrehill@cpst.org