21st Century Statistics on the STEM Workforce

A Report for the National Science Foundation's Workshop on
the Science of Science Metrics, December 2-3, 2010

R. A. Ellis

Ellis Research Services
Carlisle, Pennsylvania

November, 2010

Introduction

The coordinators of this workshop have asked me to review available data on people in U.S. science and technology. This is a mainly a matter of demography. Sources like biographies or even novels can be useful for this work, but no examination that fails to consider simple demographic facts is likely to adequately describe the entire STEM workforce. To assess data on people in these professions ("Science, Technology, Engineering, Math," for those who may be looking at this report who haven't seen the term "STEM" before), one can begin by establishing standards for good sources. No one source of STEM workforce data addresses all of the questions that the users of such information have posed. We do have at least one source that works well for most of them.

For those seeking information on specific STEM occupations, professional societies can be very good sources of data. Large associations like IEEE-USA and the American Chemical Society have impressive statistical archives, including excellent time series, on their own people. These sources meet good academic standards. STEM has many of these associations, and while they vary considerably in their propensities to collect data, their membership records alone can be valuable, as can information on their sections (for example, IEEE is organized into nearly 40 sub-associations, such as its Computer Society). Other large professional organizations include ACM (the Association for Computing Machinery) and major engineering groups like ASME (the American Society of Mechanical Engineers), ASCE (the American Society of Civil Engineers), and NSPE (the National Society of Professional Engineers, an organization for those who have become licensed PEs). The most recent *Directory of Engineering and Scientific Societies* includes well over two thousand entries (copies may still be available from its publisher, AAES (the American Association of Engineering Societies at http://www.aaes.org). These sources may yield good information on particular kinds of STEM people. However, comparisons of different STEM occupations or measures of trends in the entire STEM workforce require more general sources.

1. Issues of coverage and content

There are many requirements for good sources of statistical data on the entire STEM workforce. Sources must cover the entire nation and be comparable to national demographic, economic, and occupational data in their coverage of employment and employers, geography, and details on specific STEM professions. For all practical purposes, this immediately eliminates all sources of data other than the federal ones, along, perhaps, with international comparisons using sources like UNESCO, the Organisation for Economic Cooperation and Development (OECD), and the occasional reports on world economic regions done by staff at the National Science Foundation (NSF). No one federal source of data on people in STEM professions can cover all of the topics that users of these data require, but in combination they have made coverage of the STEM workforce feasible if not always entirely satisfactory.

Some major STEM professions such as chemistry and civil engineering have been covered by the U.S. decennial census for well over a century, but the pace of change has accelerated far beyond the original 19th century conditions. *Annual* coverage of up to at least 50 Census occupational groups is now a must if one aims to make full use of current sources. Note that most STEM workforce reporting does not presently include medicine (other than medical science), or elementary and secondary science teachers, or other fields that may include appropriate participants. Some sources (in particular, NSF) ignore all persons without at least a bachelor's degree, although many such people are in the STEM workforce. Supporting office staff like secretaries should *not* be counted as STEM people, although some industry-based reports have included them.

Some statistical indicators such as employment and unemployment trends lend themselves to quarterly tracking, as in the original editions of the publication *Engineers* done during the 1990's for AAES and its Engineering Workforce Commission (EWC). The importance of tracking trends means that although one-shot surveys can be useful, measurement of trends requires consistent time series. This also means that it is essential to be able to adjust to the revisions and additions that even the most consistent sources must put into place from time to time, because work itself changes. At one time the most common occupations in the USA had to do with the care of horses. That is no longer the case, and STEM is an arena where keeping occupational definitions up to date is especially critical. If there are weaknesses in the current federal sources of STEM data, this is one of them. For example, there is no single Census occupational category for systems engineers, and the biggest single set of engineers tracked by current federal databases is the one for "Miscellaneous Others." Other parts of STEM workforce statistics have the same problem. One should not envy the people at the Census Bureau and other agencies who have to deal with this dilemma. They completed a major upgrade of the occupational codes not long ago and already it's out of date. All is not lost for those interested in systems engineering as a profession, by the way. There's an international society for the field itself, and the electrical, mechanical and industrial engineering societies all have systems engineering sections, as may some other groups. In short, it's possible to at least point to subsets of STEM data that include the systems specialists.

The most basic source of demographic data on people in the STEM professions in the USA, then, is the federal government. No single source within that complex covers everything that people want to know. In particular, the only complete source of enrollment and degree statistics is the **National Center for Education Statistics** (for a sample, see http://nces.ed.gov/das/library/tables_listings/showTable2005.asp?popup=true&tableID=6514&rt=p). For engineering enrollments and degrees, EWC and the American Society for Engineering Education (ASEE) have more details, and additional resources are provided by the Higher Education Research Institute at UCLA, which does an excellent annual survey of entering college freshmen, and the National Association of Colleges and Employers (NACE), a source of information on starting salary offers to people in many STEM professions.

Virtually all of the remaining statistics on people in STEM comes from general federal surveys that include detailed applications of the government-wide (but sometimes augmented or abbreviated) Standard Occupational Classification (SOC) system, which allows data on those in known STEM professions to be extracted and examined. Good analysts will also use data on potential comparison groups, including the set of all professions combined and the data for all persons in the labor force as a whole.

2. The major federal sources

There are several sources of federal population statistics that can be tapped for STEM purposes. Most of them fall short in one way or another but still retain appeal for particular purposes:

- The **decennial U.S. Census** remains the basic source for long time series, such as "A Half-Century Snapshot of the STEM Workforce, 1950 to 2000" (Lindsay Lowell and Mark Regets, STEM Workforce Data Project White Paper No. 1, August, 2006, available at http://cpst.org/STEM_Report.cfm).

- The **Current Population Survey (CPS),** a joint project of the Census Bureau and the Bureau of Labor Statistics (BLS), is a good source of quarterly employment data. CPS labor force data is also available in the form of annualized MORG ("Merged Outgoing Rotation Groups") datasets which combine the monthly CPS samples into considerably larger annual samples that yield statistics on as many as 50 major classes of STEM personnel and useful details for many of the largest STEM professions. Standardized tabulations of these data, including results for 2009 plus archives for earlier years, are available at a Bureau of Labor Statistics web site at ftp.bls.gov/pub/special.requests/lf. Two of the tabular series at this location are especially useful for STEM purposes: series 11 on employment and series 39 on earned income. Both of these include results for all of the detailed SOC occupational categories. These data, along with some additional material on women and minorities maintained by BLS staff, were the major source for the *STEM Workforce Data Project,* a large undertaking sponsored by the Alfred P. Sloan Foundation with reports, white papers, and data archives also available at http://cpst.org/STEM_Report.cfm.

In the past it has been necessary to manually pull out data for STEM people from the much larger CPS tabulations. This chore can now be automated by using new online resources for both access and analysis of Census and other public databases, such as the **IPUMS systems** maintained by the University of Minnesota at http://usa.ipums.org/usa/. IPUMS now provides both access to many years of CPS and other federal databases and very good online tools for users. "IPUMS" is an acronym for "Integrated Public Use Microdata Series." The IPUMS-USA web site supports access to current and historical Census data, "integrated over time and across samples." Datasets can be downloaded for local use or explored online with the excellent UC-Berkeley Survey Data and Analysis (SDA) system.

Web sites like this one have changed the whole nature of STEM workforce data analysis. Online access and good supporting software have drastically reduced the costs of such work while also allowing many more users of this information to track STEM professions themselves, without any dependence on others. Online results from the IPUMS system are provided as downloadable HTML files that can be copied directly into spreadsheet software. Most people who can deal with spreadsheets should be able to also deal with the Berkeley package. Organizations with their own DBMS (data base management systems) capabilities may prefer to download whole Census datasets or smaller STEM-related extracts from those datasets; the downloads will include supporting scripts for users of common survey DBMS packages such as SPSS. Such users should be aware that local use will require the provision of services that the online analysis system does automatically, such as weighting the sampled data. Also note that an open source alternative to SPSS has become available: GNU-PSPP (for more information, see http://www.gnu.org/software/pspp/).

- For many years, the Bureau of Labor Statistics has generated a special tabulation known as the **Industry-Occupation Employment Matrix.** Intended mainly as support for the biennial BLS *Occupational Outlook Handbook,* this enormous tabulation (over 180,000 cells!) allows one to examine the distribution of STEM workers among different types of employers. Like the annualized CPS datasets, it's no longer necessary to do extensive manual work in order to obtain this type of data for the STEM professions; the same IPUMS resource cited above will allow users to generate similar results for the STEM professions online. See below for more information. The *Occupational Outlook Handbook* itself is an excellent source of basic information on all of the STEM professions.

- The National Science Foundation's **SESTAT** system for science and engineering statistics probably exceeds all other sources in its ability to support very detailed studies of STEM people. For example, it includes data on student minor fields of study as well as their majors. Constructed from NSF's long-standing National Survey of College Graduates (NSCG), the National Survey of Recent College Graduates (NSRCG), and the annual Survey of Doctoral Recipients (SDR), integrated SESTAT datasets for 1993 (the initial year), 1995, 1997 and 1999 have been accessed at http://sestat.nsf.gov/datatool (a CD-ROM is also available for data from these years). Additional datasets exist for 2003 and 2006. Some users have acquired these for local access, but this requires a license, typically given to academic institutions. Dr. Paula Stephan of the Andrew Young School of Policy Studies at Georgia State University is an excellent example of what such a license holder can do with SESTAT and its components; see, for example, "Job Market Effects on Scientific Productivity," available at www.cso.edu/upload/PDF_rencontres/SEM_ES_Paula-Stephan.pdf. A list of additional papers done by Dr. Stephan for the Labor Studies group at the National Bureau of Economic Research (NBER) is at http://www.nber.org/authors/paula_stephan. There seems to be a STEM workforce working group at NBER but a search of the organization's web site does not locate a page that references it. Searches of the Bureau's education and labor studies divisions do identify other appropriate publications.

  SESTAT's usefulness is more limited if one wishes to track current trends. SESTAT has not included coverage of occupations like computer programming, despite the fact that the federal occupational coding system's current revision does include this field in its coverage of STEM professions, and the system's exclusion of people without at least a bachelor's degree inhibits its coverage of both the programmers and many engineering and science technicians and research assistants. In addition, during the 1990s SESTAT was reasonably accessible and provided fairly timely data, but since then its production has been more erratic. The NSCG was not conducted in 2001 and so the system lacks a comparable dataset for that year. More recently other problems with data on race and ethnicity have been identified in the SESTAT results for 2003 and 2006. A new round of data collection was begun in Fall 2008 but those data have not yet been made available, and recent attempts to use the online SESTAT "data tool" have returned a blank page.

- A brand new source of very good data on STEM workers has become available. The **American Community Survey** (ACS, not to be confused with the American Chemical Society) is a new product of the Census Bureau. Its results appear annually and relatively rapidly; data for 2009 became available this year and were available online, with very good supporting analytic tools, in November at the same IPUMS site noted above. The ACS includes details on employment, income, citizenship, age, gender, ethnicity, highest degrees, family structures and housing, employer characteristics that are comparable to the BLS Industry-Occupation Matrix, geographic locations, and more, and the IPUMS online analysis tools are all that most users will need to tap this source (those with their own in-house resources can also download entire ACS or other Census datasets or smaller extracts). Work on the ACS began in the 1990s but full compliance with all of its sampling goals was not reached until 2006; the 2005 version comes close enough to be used for STEM workforce purposes (it does not cover people in group housing facilities, such as college dormitories, nursing homes, and prisons).

  The ACS uses a one percent sample of the entire U.S. population, making it one of the largest sample surveys in the world. A single year's dataset includes over three million cases, more than enough to enable generation of detailed data for every one of the fields covered in earlier statistics on people in STEM. For really detailed work, three year merged ACS samples triple this yield, and five-year sets will begin to appear early in 2011. Nothing like this has ever been available before, and I am convinced that as long as it continues to be available, the ACS will be the main source of data on trends in the STEM workforce. There is also a PRCS for Puerto Rico.

3. An example of an ACS-based STEM workforce data archive

Since the summer of 2009, I have been working on a large archive of STEM workforce statistics based on the American Community Survey, for the U.S. Department of Energy's Office of Science, with coordination provided by Oak Ridge Associated Universities (ORAU). Here is a list of some of the details in this source which have not been seen before in any set of annual STEM time series data:

- Employment, unemployment, and the labor force: includes data on all of the STEM occupations on these three measures for 2005, 2006, 2007, and 2008, and data for 2009 is available to be added now. Also includes data from the three-year merged 2005-2007 ACS surveys, on persons not in the labor force who still associate themselves with STEM occupations; these persons include long-term unemployed as well as retirees, people who have returned to school, those caring for families, etc. For the three-year period centered on 2006, there were over 800,000 of these additional experienced STEM professionals, and over 100,000 of them indicated an interest in returning to work.

- Geographic distributions: the archive includes a complete division of the STEM occupations among the 50 states and the District of Columbia. It also includes a division of these occupations among a set of 35 major metropolitan employment markets plus all those in any other Metropolitan Statistical Area (MSA), those outside any MSA, and the nation as a whole. These 35 metro areas account for 60 percent of the entire U.S. STEM workforce and for 70 percent of all of those in any urban area. In a few of these regions -- the San Francisco Bay area, Washington, D.C., and Raleigh-Durham, North Carolina -- STEM workers account for more than 10 percent of the entire labor force, and other areas come close to this threshold, including Austin, Boston, and Seattle.

- Industry employment sectors: the archive includes a division of the STEM occupations among the major industry groups found in common condensed summaries of the NAICS (North American Industry Classification System) categories. It also includes a division of the STEM occupations by a set of detailed NAICS categories that include every employer group with at least 100,000 STEM workers. Many STEM-intensive industry employment sectors are obvious, such as architectural and engineering services, construction, many types of manufacturers, and more. A close look at all of the possible detailed employer classes identifies other centers of STEM employment, including concentrations of people in education; finance, insurance and real estate; health and human services; information and communications services; and several major categories of public administration.

- Characteristics of STEM workers: the archive contains five large tabulations providing divisions of all of the STEM occupational groups by sex, age, ethnicity, citizenship, and highest degree. It also includes a collection of 62 separate tabulations of gender by age (in five-year groups) for 49 core STEM occupational groups plus 11 summary sets and results for the entire U.S. labor force. The contrast between the STEM professions as a whole and the general workforce is striking, not just for the reduced presence of women but also for the very different age structures of these fields, a fact that will have major implications for the supply of new talent. Interesting details abound, such as a surge of young women into the medical sciences, a demanding profession characterized by exceptionally high levels of people with doctorate degrees. Women are also showing up in the ranks of the most experienced people in many of these fields.

- Measures of demand: the archive contains a tabulation, for all of the STEM occupational groups and their summary sets, of two major measures of changes in demand from one year to the next: trends in employment and trends in earned income (standardized in constant dollars so that results reflect actual variations in purchasing power). Results appear as deviations, either positive or negative, from the overall trends for the American labor force as a whole. When plotted on a grid with one axis for employment and the other for compensation, with the zero points at the intersection of these two lines of deviation -- that is, at the center of the graph -- these indices show whether a given occupational group is doing better or worse than the labor force as a whole. Professions in the upper half of such a chart, assuming that the x axis is used for employment and the y axis is used for compensation, will have been doing better than average in obtaining real gains in purchasing power, while those in the lower half of the chart will be doing worse. Fields in the right half of the diagram will be doing better than average in employment gains, while those on the left will not be keeping up with general trends. Results suggest that gains in pay may indeed work as leading indicators of gains in employment, just as classical economists would expect. Results also indicate that as time passes, further compensation gains may not be required to maintain growth in employment. I do not claim to be a qualified labor economist and will be most interested in any reactions to these data from those who are. No regularly produced demand measures have been available for STEM professions since the cessation of production many years ago of the privately generated High Tech Recruitment Index, which was based on counts of employment recruitment advertisements in newspapers.

4. Weaknesses of STEM workforce data

In general, weaknesses of the data in the American Community Survey are common across most if not all of the federal sources of information on STEM professions. As noted above, rapid shifts in specializations now mean that the largest group of engineers reported by most federal sources of such data is the set of miscellaneous engineers not reported in any other category, and engineering is not the only STEM occupational group with this problem. Other general problems are common across most sources of federal STEM statistics, in particular the loss of an earlier capacity to include college and university professors with STEM specialties in databases used to generate statistics on the entire STEM workforce. It is probably possible to use other data sources to obtain such statistics as the number of postsecondary teachers of physics and astronomy, a figure that will help to correct for serious undercounting for that profession. Some STEM fields comprised mostly of academics, such as history and political science, disappear altogether from statistics on STEM occupations because of this problem, which did not exist under the version of the U.S. Standard Occupational Classification (SOC) coding system that was in place before the latest major revision. The Sloan-sponsored STEM Workforce Data Project, noted above, includes a white paper with details on this and a number of other weaknesses of current federal STEM statistics. It also notes that despite those weaknesses, the current system is very useful.

For many years most archives of statistics on the STEM workforce have adhered to Census conventions for organizing the list of professional categories that can be associated with this part of the U.S. labor force, because this practice could help other consumers of the data find and use the same sources. This practice is no longer needed, because online systems have replaced dependence on more obscure resources requiring much more effort to access and use. STEM data archives can now adopt a few revisions of the Census practices that improve the fit of these data to STEM specialties. For example, earlier presentations of the Census STEM occupational groups place mathematics and its associated fields within the set of IT (information technology) professions, but math is the bedrock of all science and should be counted as a basic discipline rather than as a part of any particular set of applications. If the “M” in “STEM” should stand for anything, it should be for medicine, which is the largest available example of an applied scientific realm like engineering that (except for coverage of medical scientists) is not otherwise treated in most of the existing work on this topic, probably mainly because of the sheer cost of covering all of it.

In addition, my own work on these data continues to follow other conventions established for the Sloan project cited above, which call for the inclusion of three kinds of STEM managers, the exclusion of architects (who have been covered in similar work sponsored for many years by the National Endowment for the Arts), and the inclusion of a relatively small set of sales engineers in the group of miscellaneous engineers (this group, classed as a sales occupation, includes trained professionals whose work entails the selection of appropriate technology for a client's specific situation; over 70 percent of these workers have a bachelor's degree or better). These steps allow users of the new ACS-based STEM data archive to use matching data from the Sloan projects for up to 20 earlier years.

Richard Ellis

Ellis Research Services
125 Walnut Street
Carlisle, PA 17013

Satellite office: Mt. Holly Springs, PA

raellis@earthlink.net

(717) 386-7931 (cell)
(717) 218-9818 (voice and fax)