

Report on Study Visit to Washington DC by Judith Glover for the
Women and Science Unit, Science and Society Directorate, Research
Directorate General, European Commission, January 2001

Judith Glover
School of Sociology and Social Policy
University of Surrey Roehampton
UK

April, 2001

Contents of Report

	Page number
1. Acknowledgments	2
2. Executive Summary	3
3. Recommendations	4
4. Examples of US data sets relating to women and scientific education/employment	6
4.1 National Science Foundation data	6
4.2 National Center for Education Statistics data	7
4.3 US Census Bureau data	8
4.4 Other data of interest	8
5. Use of data	9
5.1 Publications of interest to a range of data users	9
5.2 Data access	11
5.3 Data dissemination	12
5.4 Using data as a basis for status reports	13
6. Issues arising	13
6.1 Commercial sector data	13
6.2 The qualitative approach to indicators	13
6.3 Harmonisation issues	14
Appendix 1: SESTAT variables	17
Appendix 2: List of experts visited	19

1. Acknowledgments

I would like to thank the many people in Washington DC who gave of their time to explain the collection and use of data relating to women and scientific education/employment in the US. These experts are listed in Appendix 2 of this report, but not all of their colleagues and assistants have been named. I would especially like to thank Shirley Malcom, Director of Education at the American Association for the Advancement of Science, who arranged for Virginia Van Horne to set up a timetable of visits to experts who could throw light on the data relating to women and scientific education/employment.

2. Executive Summary

This study visit, funded by the Women and Science Unit, Science and Society Directorate, Research DG of the European Commission, had two aims: to find out about a) the range of statistics which are available in the USA on women and scientific education/employment and b) the use to which the statistics are put.

The statistics available from government agencies in the USA are plentiful and detailed. Much importance is attached to the production of high-quality statistics and the resources required for this are undoubtedly considerable. Longitudinal data and highly detailed cross-sectional data are available, covering both education and employment. Surveys of the science and technology workforce are carried out mainly by the National Science Foundation. Other more general surveys are also useful for secondary analysts who are seeking to explore the issue of the science and technology workforce. Whilst data gatherers insist that there is a clear distinction between gathering data and the use to which they are put, there is evidence that influential institutions make good use of the available data in order to make the case at the highest level for the development of policies relating to the science and technology workforce.

The importance attached to data gathering in the USA, as well as the breadth and depth of available data, could serve as an example of good practice for the European Union countries. Several recommendations are made (see over).

3. Recommendations

3.1 Data collection

3.1.1 To consider the practice of oversampling in order to investigate the science and technology population (both in education and employment), for example as an add-on to an existing data set, such as the national labour force surveys.

3.1.2 To explore the possibility of collecting data in developing fields of science and technology, such as the biosciences and computing. The lack of data in these areas presents an opportunity for different countries to work together in order to ensure harmonisation from the start of data collection. Data harmonisation could well be achieved efficiently in less established scientific fields, such as the biosciences and computing, where custom and practice is still at a fluid stage.

3.1.3 To consider using countries which have developed rather little in the way of data on scientific education and employment as 'trials' in terms of exploring the gathering of harmonised, but also detailed, data. These could serve as examples of good practice which could be instructive for those countries which have a much longer history of data collection..

3.1.4 To develop indicators which are qualitative as well as quantitative. The qualitative approach is likely to be particularly important in the investigation of best practice (for example the factors which appear to have led to success in a particular area).

3.1.5 To encourage countries to co-operate with their national Census bureaux in the collection of data and the further exploitation of existing data. The advantage of this is that Census bureau personnel have experience of surveying very large national populations, something which is necessary for the production of adequate statistics on small populations within the national populations. Furthermore, the example of the use by the National Science Foundation of Census data to establish a sample of scientists and engineers who are then the subject of a separate survey could well be of considerable interest to many countries.

3.1.6 To encourage co-operation with supranational organisations such as OECD. If countries are able to conduct add-on surveys to their existing surveys, such as the Census or the national labour force surveys, it is important that harmonisation issues are considered. The expertise of Eurostat would be essential here and that of OECD very useful.

3.1.7 To encourage professional bodies to co-operate with one another in their collection of data on their members, by developing a common set of questions. This would enable cross-disciplinary analyses to be made.

3.2 Data Use

3.2.1 To consider the feasibility of establishing a resource which brings together a range of data sources, such as the CPST's *Professional Women and Minorities: A*

Total Human Resources Data Compendium. To be noted, however, that the resources needed to bring this level of detail together and to maintain the resource are undoubtedly considerable.

3.2.2 To consider producing regular reports such as the National Science Foundation's *Women, Minorities and Persons with Disabilities in Science and Engineering*, which successfully manage to address both a generalist and a specialist audience.

3.2.3 To consider the establishment of an on-line data resource such as the SESTAT database. To note, however, that considerable resources would be required to set it up and maintain it adequately, as well as to provide support for users, since the amount of statistical knowledge required to use such a database meaningfully should not be underestimated.

3.2.4 To bear in mind the media used for the dissemination of information - paper or online. It would be regrettable if the move towards online dissemination resulted in access being compromised.

3.3 Substantive Issues

3.3.1 Bearing in mind the decreasing representation of women in Information Technology education and employment in the USA and the UK, to investigate this issue in EU member states.

4. Examples of US data relating to women and scientific education/ employment.

Three data gatherers are singled out here for particular attention. They are the National Science Foundation, the National Center for Education Statistics at the Department of Education and the US Census Bureau.

4.1 National Science Foundation Data

The federally funded National Science Foundation (NSF) is an example of good practice in terms of data collection and data dissemination on the issue of the scientific workforce. The section of the NSF which is particularly relevant to the issue of women in science is the Division of Science Resources Studies (SRS). Its mission is 'to provide policymakers, researchers and other decision-makers with high quality data and analyses for making informed decisions about the nation's science, engineering and technology enterprise'. It fulfils this mission by designing, sponsoring and conducting surveys; analysing and synthesising data, both national and international; developing indicators; preparing and disseminating reports for its users. Its users are: government officials; members of Congress; planners; the news media; social, behavioural and economic scientists. Its statistical reports and other material are available at <http://www.nsf.gov/sbe/srs/>.¹

Of particular interest to the questions which have been the focus of much of Research DG's Women and Science Unit's work is a range of major surveys which cover both education and employment in the sciences. These are the Survey of Earned Doctorates, the Survey of Doctorate Recipients, the National Survey of College Graduates, the National Survey of Recent College Graduates and the Survey of Graduate Students and Postdoctorates in Science and Engineering.

A primary method of access to most of these data sets is the SESTAT system (<http://srsstats.sbe.nsf.gov/>). This is an integrated on-line system (see Section 5.2 below) which combines three of these surveys (the Survey of Doctorate Recipients, the National Survey of College Graduates and the National Survey of Recent College Graduates). The system contains data from 1993 up to 1997 and the 1999 data are shortly to be added. The 1993 surveys include responses from about 200,000 individuals, representing an S&E population of 11.6 million. The 1995 surveys include responses from about 105,000 individuals, grossed up to 12.0 million scientists and engineers. The 1997 surveys include responses from about 100,000 individuals grossed up to 12.6 million scientists and engineers. The 1993 National Survey of College Graduates is a decennial baseline survey which covered the S&E and non-S&E population with a bachelor's or higher degree -- grossed up to 29 million people. The bringing together of different time points means that time series are possible. This is an example of the way in which cross-sectional data, if sustained over time, can be very useful.²

The variables included in the SESTAT integrated database are in Appendix 1. These represent only a small proportion of the questions asked in the surveys. However, they would answer most of the questions to which the Women in Science Unit is interested in finding answers. It will be clear, however, that this level of detail is only sustainable, if there is a large sample size and considerable resources to establish and sustain the database.

Methods such as oversampling of small sub-populations may therefore need to be considered by smaller countries. This is a technique where small populations of particular theoretical or policy interest can be deliberately oversampled, so that sophisticated statistical techniques can be used to analyse the data without fear of sampling error and/or lack of representativeness, due to small sample size. This is also a particularly important technique when fine distinctions are needed within small populations. By the same token, large groups can also be undersampled. For example, in almost all industrialised countries, including the US, there are now very large numbers of business studies students and small numbers of science and technology students. For some purposes, therefore, it makes sense to undersample the business studies students and oversample the science and technology students.

4.2 National Center for Education Statistics data

The National Center for Education Statistics (NCES) organises several surveys which are of considerable potential interest to the issue of women in scientific education, and these data provides also some information on the first destinations of scientifically qualified women (<http://www.nces.ed.gov>). No special surveys are carried out which focus specifically on this issue, but secondary analysis has the potential to single out this sub-population. Because the data sets are so large and often very detailed, the analysis of sub-populations is relatively unproblematic. Many of the data sets are longitudinal, useful for the study of attrition from scientific education, something which the Women in Science Unit at Research DG is particularly interested in. For example, the National Educational Longitudinal Study is an NCES study which has followed several cohorts over time. The cohort which graduated from High School in 1972 was followed for fourteen years, yielding data on college, university and the job market. The sample size was originally around 25,000 and this had decreased by the time of the final sweep to 12,000, when the cohort was aged around 32. This reduction in sample size was due to budget constraints and oversampling of small sub-populations was then used. The response rate was 90% or higher each time. Another cohort study is the National Post-Secondary Student Aid Study, which follows secondary school students over a period of around five years. This covers the period through to university. The sample sizes are between 40,000 and 60,000.

Other NCES data sets of interest are Beginning Postsecondary Students; Baccalaureate and Beyond; High School and Beyond; National Longitudinal Study of the High School Class of 1972; National Study of Postsecondary Faculty (which provides evidence on university teachers and researchers). The Integrated Postsecondary Education Data System (IPEDS) is a system of surveys designed to collect data from all primary providers of postsecondary education. IPEDS is a single, comprehensive system designed to encompass all institutions and educational organizations whose primary purpose is to provide postsecondary education. The IPEDS system is built around a series of interrelated surveys to collect institution-level data in such areas as enrolment, degree completions (Masters and doctorates) and faculty employment. It has the advantage of having a very high level of detail in the fields of study. The findings are publicly available on the Internet (<http://www.nces.ed.gov/ipeds/>).

The Third International Mathematics and Science Study (TIMSS) study should also be mentioned in this section. Since 1960, this Department of Education study has compared the mathematical achievement of secondary school students in a wide range of countries (<http://www.nces.ed.gov/timss/>). Although, strictly speaking, it is not of direct interest to the Women in Science brief, its methodology and approach to cross-national work has potential interest for the work of the Helsinki Group. These points are made in more detail below, in section 6.2.

4.3 US Census Bureau Data

When there is a need to research small sub-populations, countries are often faced with the problem of not having sample surveys which are of a sufficient size to research these small populations adequately. The issue of women and scientific education/employment is a clear example of this (Glover and Bebbington, 2000). Use of census based surveys may therefore be worth considering for countries which have much smaller national populations than the US. The US Census Bureau's Survey of Income and Programme Participation (SIPP), which is based on the population of the 10-year Census, is a longitudinal survey which follows a panel of households every three years. The aim of the survey is to examine participation in government support programmes, welfare, unemployment insurance and other programmes. It includes the highest degree, the field of degree and has a sex variable. In addition it includes current occupation and family situation information. Because of its large N, such a study has the potential to study small subpopulations, such as scientists. A further issue of interest is the Census Bureau's practice of attaching different topics of special interest to the population surveys which are carried out frequently. These topics are selected by government departments.

Of particular interest is the link between census data and one of the NSF surveys, the National Survey of College Graduates. This is a survey based on a sample of people identified from the decennial census in either science or engineering occupations or who have at least a bachelor's degree in science or engineering. The survey is combined with the Survey of Recent College Graduates and Survey of Doctorate Recipients to form the SESTAT database (see Section 4.1 above).

4.4 Other data of interest

Many professional bodies collect detailed data on their membership, examples being the American Chemical Society and the American Institute of Physics. The American Chemical Society has two regular surveys, one on the new entrants and the other on those already in the field. Detailed data are collected on education, details of jobs (including private/public sectors) and salaries. Every five years, the entire membership (around 180,000 members) is surveyed. A particularly interesting feature is that there is a good level of co-operation between some professional bodies in terms of survey design, such that members of different organisations are asked a common set of questions. This is a particularly important point since it means that comparisons can be made on a 'like for like' basis between disciplines. An important factor here is that there needs to be an umbrella body which will coordinate the efforts of different professional bodies. In this case, this coordination was supplied by the Commission for Professionals in Science and Technology (CPST).

However, professional body data are not representative of all people with science qualifications and tend to be much more complete for those with postgraduate degrees than for those with first degrees only. Furthermore, not all disciplines are well-organised in terms of having professional bodies. A particular issue here is that there is rather little information on the biosciences, a field with a high representation of women, since no one professional body covers this field. This is a similar issue for the field of computing, a relatively new area. A problem relating to computing is that there is a lack of consensus as regards the definitions of who should be included and also about how jobs should be defined.

The Department of Labor's Bureau of Labor Statistics also produces longitudinal data (NLS, National Longitudinal Surveys). Whilst these are not specifically aimed at the science and technology workforce, they have some potential for secondary analysts who are seeking to explore questions relating to sub-populations (<http://www.bls.gov/nlsdata.htm>).

5. Use of data

It was emphasised by the data gathering organisations that a distinction needs to be made between them and organisations which make use of those data, such as other agencies within the government machine. An example of the latter is the Women's Bureau at the Department of Labor which uses statistics produced by the Bureau for Labor Statistics, housed in the Department of Labor. The National Academy of Science (www.nas.edu) is an example of a high profile organisation which makes use of the quality data produced by organisations such as the NSF and the NCES. Its role is to advise Congress to whom it produces an Annual Report. The NAS is thus in a strong position to influence policymakers. A further example of the use of data is that made by the National Science and Technology Council. This cabinet-level council was established by President Clinton in 1993 and is the principal means for the President to coordinate science, space and technology policies across the federal government. It is chaired by the President and membership consists of the Vice President, the Cabinet Secretaries and Agency Heads with significant S&T responsibilities. Its reports make particular use of the data gathered by NSF (National Science and Technology Council, 2000).

Another aspect of data use involves the establishment of links between data users and data gatherers. The Department of Education's National Center for Education Statistics places particular importance on this. It has set up 'data user' groups and one function of these is to discuss the topics selected by the Department of Education for future development and to comment on and possibly add to the Department's list of topics.

5.1 Publications of interest to a range of data users

One example of the use of statistics is the Report of the Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development (CAWMSET Report, National Science Foundation, 2000). Here, the National Science Foundation was charged by Congress to report on the

status of women, as well as other social groups who were underrepresented in scientific education and employment. This report took the 'pipeline' perspective, combing the available data at the different phases through which people go to reach a career in science. Data were available to shed light on pre-college education, with a particular focus on mathematics and science; on access to scientific Higher Education; on attrition from scientific higher education; on achievement in scientific higher education. Trend data were available and these are particularly interesting in relation to Information Technology, showing that the percentage of women earning computer science degrees has dropped steadily since 1984 (CAWMSET report, p 40). The report also shows women's representation in different sciences, with distinctions made between life sciences, physical sciences, engineering and computer science. Information was also available on salary differentials, on women's representation in industry and government and their representation in high-level corporate positions. The Congressional Commission aims, amongst other things, to monitor progress through ongoing data compilation and analysis, and it has established targets which relate the representation of women and other groups in science and technology to their representation in the working population. The CAWMSET Report is an indication of the breadth of data which are available in the US. It has a clear aim, which is to make the 'business case' for introducing more diversity into scientific employment.

In the area of education, the National Center for Education Statistics at the Department of Education has used cohort data to follow the trajectories of people with science degrees (National Center for Education Statistics, 1996). The Department of Education was particularly interested in the supply of teachers and wanted to know how many of the science graduates went into teaching as an early destination. This area is obviously highly policy relevant.

Information of a more in-depth nature is available from the National Science Foundation's two-yearly reports *Women, Minorities and Persons with Disabilities in Science and Engineering*. These reports, mandated by Congress since 1982, provide an invaluable time-series of a range of issues relating to undergraduate enrolment and achievement, postgraduate enrolment and achievement and science and engineering employment. They are particularly useful for their accessible presentation of text and graphics, combined with detailed appendix tables. Thus, they successfully address two audiences: the data user with a general interest will find the text and graphics particularly useful and the specialist researcher will find the appendices, with their large amount of detail and frequently arranged as time series, extremely helpful. To illustrate this latter point, take for example Appendix Table 4-11 (National Science Foundation, 2000: 188). This shows women as a percentage of all doctoral recipients in an annual time series between 1966 and 1997. It is disaggregated by 'major field group': engineering, physical sciences, earth, atmospheric and ocean sciences, mathematics, computer science, biological/agricultural sciences, psychology, social sciences and non-science and engineering. A further example of the available detail is Appendix Table 5-1 (National Science Foundation, 2000: 210). This shows, for one time point (1997) the presence in the labour force of different types of women and men scientists by type of degree (first degree, masters, doctorate). There are 31 different types of scientists. For example, taking physical and related scientists, this is sub-divided into chemists; earth/ geology/ oceanographers; physicists and astronomers; other physical scientists; post secondary physical science teachers. Typically, when this level of detail is shown, only one time point is shown, for

reasons of space and complexity. However, identical tables are usually available in previous reports. For example, the 1996 Report shows an table, Appendix Table 5-1, which is identical to that in the 2000 report, using the same categories and conducting the same analysis. Both tables use raw numbers, rather than percentages. Thus, data users can make calculations and create their own categories for their own purposes at a high level of complexity. Furthermore, they are able to make comparisons over time, by reference to different volumes, if necessary.

An example of the production of highly detailed data from a range of sources is the Compendium produced by the Commission on Professionals in Science and Technology (CPST), entitled *Professional Women and Minorities: A Total Human Resources Data Compendium*. This publication uses the range of US data which are available on women in scientific education and employment and is now in its 13th edition (CPST, 2000). The data include reports from government agencies, professional associations and interest groups. In common with the National Science Foundation's *Women, Minorities and Persons with Disabilities*, its principal advantages are that it shows time series and that it recognises the importance of distinguishing between the different sciences. Thus, for example, it contains details of women's doctorates in the earth sciences/environmental sciences/marine sciences, disaggregated into 28 subfields, over a period of 38 years (1960 to 1998). These data are from the National Science Foundation's Survey of Earned Doctorates. Thus the Compendium uses data which are already available elsewhere, although not at this level of detail, at least in the public access files. The Compendium also contains data from less widely available sources such as the Engineering Workforce Commission and from professional bodies such as the American Chemical Society and the American Mathematical Society.

As Shirley Malcom, Director of Education at the American Association for the Advancement of Science, says in her foreword to the 2000 Compendium, the facts about our human resources landscape are crucial in order to change "hearts, politics, programmes and practices". In particular, such a publication is essential in terms of documenting past trends and hence predicting future change. Dr Malcom emphasises the importance of such a document in order for arguments to be made by a variety of people who are in a position to influence policy. A particularly important aspect is that the Compendium does not hide the methodological background to the different surveys which are used and indeed draws attention to different definitions, where appropriate. Furthermore, its tables often contain large numbers of footnotes which direct users towards an informed use of the data. This is the sort of 'working' publication which forms the crucial basis for user-friendly reports.

5.2 Data Access

The openness of data access is striking in the US. This approach was justified in the following way: if data are openly available, this allows both specialist users and the general public to verify the data and they will hopefully feel greater confidence in the organisations which are producing the data, because secrets are not being kept from them.

Much of the data is available publicly. Access can be in the form of reports, CDs or Internet access to databases. Typically, public access data are at a low level of detail. Bona fide researchers can apply to most agencies to have more detailed 'restricted' data made available. Academic access can also be on an institutional level, such that universities can apply for an access agreement which covers the use of the data by any academics employed by that institution.

Undoubtedly, a move away from publications in paper form was a consistent trend in the organisations visited. Whether this will decrease accessibility is an issue which requires some thought. Whilst complex tables can be viewed with some ease on paper, this is much more difficult on a computer screen. Furthermore, many printers are not able to print tables as they appear on the screen. The move away from paper is largely for resource reasons, but it does raise questions about data access.

There is a move towards putting databases on the Internet and this is exemplified by the National Science Foundation's WebCaspar on-line database and the SESTAT database which deals specifically with science and technology. The National Science Foundation's SESTAT system has already been mentioned earlier (Section 1.1). As well as on-line access (<http://srsstats.sbe.nsf.gov>), SESTAT data are also available in CD format which contains cross-sectional, time series and longitudinal analysis with data formats in SAS or flat file. This is particularly useful for researchers who carry out secondary analysis. SESTAT enables users to specify their own query from a very wide list of variables which are brought together from different surveys. Some statistical knowledge is necessary in order to create meaningful tables. The National Science Foundation also has a database, WebCASPAR, which provides access to a wide range of data sources relating to science and engineering at academic institutions (<http://caspar.nsf.gov/webcaspar>). Information can be downloaded into tabular form or as spreadsheets. However, the challenges involved in making such databases user-friendly to a range of users with variable statistical skills and knowledge are considerable.

The Census Bureau also makes available its data for public use (PUMS data, <http://www.Census.gov>). The Bureau of Labor Statistics of the Department of Labor makes its data available on CD and on its website (<http://www.bls.gov>). As well as making data available on its website, the Department of Education's National Center for Education Statistics has a Public Use Data Analysis System (DAS) which allows users to carry out data analysis of many of its data sets. This is available on CD. Knowledge of quantitative data analysis principles and software is required to make full use of this facility, however.

5.3 Data dissemination

A point which was emphasised by several experts was that dissemination of complicated data needs to take the audience into consideration. There are several audiences, including the general public, policymakers and journalists. Accessibility of information is an important issue in the US (and of course elsewhere) and clear presentation through the use of simple graphs and charts is important for general audiences. Furthermore, a focus on successful stories rather than unsuccessful ones (in other words, a focus on best practice) was emphasised as the correct approach.

5.4 Using data as a basis for status reports

A particularly interesting use of data was mentioned by the Director of the Committee on Women in S&E at the National Research Council. She is working with her Committee Chair, who has a background in the business sector, on the production of status reports or 'report cards' of women's position in US research universities. The idea is to develop indicators for each of these universities, focusing on aspects such as the number of women scientists at different levels, the number of women science students applying and graduating and so forth. This will then be published, probably on the Internet, so that interested parties can access the data and draw their own conclusions. Thus, without explicitly ranking performance, examples of good (and less good) performance can be highlighted.

6. Issues arising

6.1 Commercial sector data

Despite the overall richness of data in the US, the acquisition of data from the business/commercial sector on employment issues was pinpointed as particularly problematic. The reasons for this are similar to the reasons which underpin the lack of data in this area in Europe: first, confidentiality issues (which usually relate to salary data) and second, the business world is not convinced that collection of anything other than baseline data is a good use of their resources. It was suggested that the confidentiality issue could be approached by recognising that the business sector is unlikely to reveal details of salaries per se, but they are more likely to be happy to show that the salary gap between groups of their employees (ethnic minorities, women) has narrowed over time. In relation to the 'burden on business' issue, one suggested way forward was to capitalise on the business sector's interest in the identification of best practice. Typically, this interest is based on establishing pointers to greater profitability; it was suggested that this aim should be emphasised in any approach to the private sector for the collection of data. However, it was suggested that some of these problems can be overcome by collecting data from individuals who work in the private sector, not from companies themselves. Also, the representation of the business/commercial sector on support/user groups was pinpointed as particularly important in order to gain early co-operation and confidence from the business/commercial sector.

6.2 The qualitative approach to indicators

The investigation of best practice is likely to require both quantitative and qualitative data. If an aim is to identify best practice in order to get more diversity into post-doctoral science and engineering, for example, it is possible to identify areas of success and to initiate studies which look at the elements which appear to have led to that success. Similarly, a qualitative approach is useful when looking at the success rate. Here, the quantitative approach to citation may not be particularly useful when looking at issues of gender. Some sort of qualitative assessment of the quality of papers needs to be made, as well as simply counting citations. This approach would accept that the citation of sources has a social aspect as well as a purely objective one,

an argument which sociology has consistently made (Law and Williams, 1982; Delamont, 1989; Rose, 1994).

6.3 Harmonisation issues

The issue of cross-national data harmonisation was discussed with experts from the Department of Education's National Center for Education Statistics, particularly in relation to the Third International Mathematics and Science Study (TIMSS). This is a survey which has compared since 1960 the mathematical achievement of secondary school students in a wide range of countries (<http://www.nces.ed.gov/timss/>). Seven main points were made in relation to cross-national studies of this sort and these points may be particularly helpful for the exercise which the Women in Science Unit at Research DG is carrying out in relation to indicators on women in scientific education/employment in member and applicant states. First, the importance of clear and detailed documentation was stressed, so that data users know exactly what has been collected. Even if the data are not exactly the same, a high level of documentation increases confidence amongst data users and minimises the opportunity for different countries (or states) to account for substantive differences in findings by reference to methodological differences in the data. Second, whilst harmonised data can be a long-term aim, it is possible for surveys to start off in an unharmonised way and then to move towards harmonisation. In this way, data collection becomes a learning process in which all countries are involved. Third, the experience of the TIMSS Survey and its predecessors showed that it is a advantageous for countries which have more and possibly better data to move on at a faster pace than countries with less data, because in so doing they can set up targets for other countries. Fourth, countries which have established good practice in terms of data collection, for example by ensuring the full commitment and involvement of their national statistics agencies, can serve as examples of good practice for those countries which are not able to set up such systems. In this way any unevenness in terms of methodology and expertise can be tackled. Fifth, the importance of having powerful national representation was stressed, in the form of a Minister of Education or a Minister of Economy. Such high-level representation would ensure support for the allocation of national resources for the collection of data or the further exploitation of existing data. Sixth, even if harmonised data are not available, national longitudinal data or a time series of national cross-sectional data allow for the establishment of trends and the investigation of best practice, which is subsequently useful to other countries. Seventh, countries with less developed statistical resources can sometimes be used to trial new approaches to the collection of harmonised data; in this way their short history of statistical development can be an advantage.

Notes

¹ The National Science Foundation fulfils many functions other than the gathering of data. From the point of view of this study visit, a particularly interesting function is the intervention projects which seek to encourage girls and women to undertake scientific education and employment. An example of this was the Programme for Gender Equity in SMET (Science, Mathematics, Engineering and Technology) whose aim is to broaden the participation of girls and young women in SMET education and careers. It does this through funding different types of projects, such as research, demonstrations, dissemination and planning grants.

² One of the data sets, the Survey of Doctorate Recipients (SDR), in the SESTAT system is longitudinal, although longitudinal data analysis cannot be performed in the SESTAT system itself. Samples of new cohorts are added to the base sample every two years. The sampling frame for the SDR is constructed from the Doctorate Records File (DRF), a historical database derived from the Survey of Earned Doctorates, an ongoing census of all U.S. doctorate recipients since 1942.

The SDR frame is restricted to two groups: (1) S&E doctorates under 76 years of age who are U.S. citizens and (2) non-U.S. citizens who plan to remain in the U.S. after they receive their degree. For the 1993 SDR, there were 568,726 from the sampling frame, 49,228 of whom were sampled. In 1995, a sample of new cohorts - those earning doctorate degrees at U.S. institutions between July 1, 1992 and June 30, 1994 - was added, and the previous sample of doctorate recipients (degrees received January 1, 1942 to June 30, 1992) was subsampled. The combined sample was about the same size as the 1993 sample. New versus old cohorts were sampled at similar rates within strata defined by demographic group, field of study, and sex. The sample design for the 1997 SDR was much like that of the 1993 and 1995 SDR. In 1997, a sample of new cohorts - those earning doctorate degrees at U.S. institutions between July 1, 1994 and June 30, 1996 - was added, and the previous sample of doctorate recipients (degrees received January 1, 1942 to June 30, 1994) was subsampled. The combined sample was about the same size as the 1995 sample. (<http://srsstats.sbe.nsf.gov/techinfo.html#1997SDRDesign>).

References

- CAWMSET Report, National Science Foundation, 2000. *Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering and Technology*. Report of the Congressional Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development. Arlington, Va: NSF.
- Commission on Professionals in Science and Technology (CPST). 2000. *Professional Women and Minorities: A Total Human Resources Data Compendium*. 13th edition. Washington: CPST
- Delamont, S. 1989. *Knowledgeable Women: Structuralism and the Reproduction of Elites*. London & New York: Routledge.
- Glover, J. and Bebbington, D. 2000. *Women and Scientific Employment: Mapping the European Data (a directory investigating data availability in EU countries)*. Brussels: European Commission Research Directorate General, Women and Science Sector.
- Law, J. and Williams, R. 1982. 'Putting Facts Together'. *Social Studies of Science* 12: 535-8.
- National Science and Technology Council (NSTC). 2000. *Ensuring a Strong US Scientific, Technical and Engineering Workforce in the 21st Century*. Washington: NSTC. Executive Office of the President of the United States
- National Science Foundation (NSF).1996. *Women, Minorities and Persons with Disabilities in Science and Engineering: 1996*. Arlington, VA. NSF 96-311.
- National Science Foundation (NSF). 2000. *Women, Minorities and Persons with Disabilities in Science and Engineering: 2000*. Arlington, VA. NSF 00-327.
- National Centre for Education Statistics (NCES). 1996. *Out of the lecture hall and into the classroom: 1992-93 College Graduates and elementary/Secondary school teaching*. Washington DC: US Department of Education, National Center for Education Statistics. NCES 96-899
- Rose, H. 1994. *Love, Power and Knowledge*. Cambridge: Polity.

Appendix 1: Variables included in the SESTAT integrated database

For the employed:

- Primary job and salary during reference week
- Full/part-time status -- including reasons for part-time
- If previously retired
- Type of employer: educational institution (by type); private for-profit; private not-for-profit; government (by level)
- If self-employed
- Supervisory responsibility, including number typically supervised directly and through subordinates
- Relationship between work and education (highest degree) -- including reasons for employment outside the degree field
- Typical work activities (in 14 categories) -- including primary and secondary work activities
- Second job -- including occupation, salary, and relationship between work and education (highest degree)

For the unemployed and those not in the labor force:

- Reasons for not working during the reference week
- When last worked
- Job last worked

Other work-related information:

Labor Force Status: Five Years Ago

- Type of employer and job
- If different from current job, reasons for changing employer or job

Professional Activities:

- Membership in professional societies and associations -- including meeting attendance
- Participation in work-related activities -- including types of activities and reasons for participation

Education:

- First degree and two most recent degrees -- level, degree field (major and minor), date awarded
- Earlier education -- date awarded high school diploma; Associate degree(s)
- Continuing education -- post-degree college courses, reasons and field of study; employer financing
- Work-related training -- types of work-related training and reasons for training activities

Other Information:

Family-related:

- Marital status
- Spouse's employment status -- if working full/part-time, technical expertise required on job
- Children living at home (and ages)
- Parents' educational attainment

Demographics:

- Citizenship status (by type)
- Age
- Race/ethnicity
- Sex
- Disability
- Country of birth

Appendix 2. List of experts visited

Joan Burrelli, Corinne Alfeld-Liro, Mary Golladay, Susan Hill, Carolyn Wardle,
Ruta Sevo,
Science Resources Studies, National Science Foundation.

Beatriz Chu Clewell
Urban Institute (Secondment to NSF to direct the CAWMSET report)

Paula Knepper, Andrew Malizio, Patrick Gonzalez, Mariann Lemke
US Department of Education , National Center for Education Statistics

Ann Mullen (Researcher on women in science)
US Department of Education, National Institute on Postsecondary Education

Jong-on Hahm,
Committee on Women in S&E, National Academies of Science

Sandra Hanson (Academic writer on women in science)
Catholic University

Harriett Harper,
US Department of Labor, Women's Bureau

Shirley Malcom,
American Association for the Advancement of Science (AAAS)

Annetta Smith,
US Census Bureau, Special Population Branch