## Robert W. Cain, National Science Foundation

#### Introduction

The National Science Foundation continually examines the needs for information on scientists and engineers, the data being collected, and the means used to collect such information. On the basis of this continuous review the Foundation has designed data collection systems, which seek to obtain the most comprehensive and reliable information for the least cost. The National Register of Scientific and Technical Personnel was a part of this system. While it was always realized that the use of the data collected through the Register mechanism was limited because of the uncertainties introduced by a relatively large non-response rate, it became increasingly apparent that the operation of the National Register was too costly. Furthermore, it became evident that the corollary objective of the Register of acting as a locator of scientific and technical personnel in times of national emergency could no longer be justified.

Therefore, the President's budget for FY 1972<sup>1</sup>/, as submitted to the Congress, recommended the "discontinuation of the National Register of Scientific and Technical Personnel in its present form." It was further recommended that funds be appropriated "to allow for the development of alternative mechanisms for obtaining required information on scientists and engineers, ---." In addition, the Foundation with the request for FY 1972 funds was to "examine new methods of making scientific and manpower analyses."

The National Register of Scientific and Technical Personnel had the following features: The information was collected from individual scientists and engineers; data could be secured directly on the personal and professional characteristics of these persons; the information on individuals could be processed with a flexibility impossible in systems collecting only aggregate data and; it provided a capability for longitudinal analysis of matched individuals.

The Register, though, was only a part of the scientific and engineering manpower data program within the Federal Government. Several other programs exist. Some of these are supported and/or directed by the Foundation; others are operated independently by other agencies with their output being used by NSF. These programs primarily involve collections of information from employers of scientists and engineers and produce summary information on the numbers employed and functions in which they engage. However, these types of collections do not provide a capability to collect detailed data on individual scientists and engineers, such as age, level of degree, work experience, or scientific specialty, primarily because in most instances, employers generally do not maintain records on such subjects and because they are usually not willing to provide anything but aggregated information. Thus, this method of collection does not

provide a capability to obtain data on matched individuals for longitudinal analyses.

## Determination of Needs for Data on Scientists and Engineers

The determination of needs for data must be based on the specific or stated and as well the anticipated requirements of policy or decisionmakers, both within and outside the Government. Such information also serves the requirements of persons desiring information for the analysis of technical manpower issues and problems. These needs for data vary by sub-group of the science and engineering population, such as level of educational attainment (doctorate and nondoctorates), sector of employment (education, Government, industry, etc.), and activity (R&D, etc.). Thus, various data elements are essential for the analysis of a given subject at a given point in time. Furthermore, a data collection system must incorporate the capability to supply needed information on a continuing basis. The mechanisms selected, however, need not provide for the continuous collection of any given data element. Rather, specific elements can be obtained at various time intervals from a given sub-group of the population with some elements collected more frequently than others.

NSF examined the data elements formerly incorporated in the Register operation from the standpoint of use and reliability. Other data elements, not previously collected through the Register, were also examined in terms of essentiality both for decisionmakers and other users. Also, reviewed were previously stated needs for information as expressed in requests of various bodies, such as the Congress, the National Science Board, the Executive Office, other NSF units, etc. These examinations indicate that two general classes of data elements can be set forth: Essential Needs--that is those elements which are needed on a more or less continuous basis from all subgroups of the science and engineering population; and Lesser Needs--those elements which might be collected less frequently or only from one or more subgroups. Essential Data Needs include: age, sex, level of degree, employment status, types of employer, program employment, work activity, Government support, field or specialty, annual salary, and postdoctorate status. Lesser Data Needs include: citizenship, student status, total professional income, work experience, and secondary school education.

#### Development of Alternative Systems

The development of alternative mechanisms to the National Register was undertaken by the National Science Foundation on the basis that these mechanisms should supply selected data on representative samples of individual scientists and engineers in the absence of the Register and be responsive to the development of an adequate, timely, and relatively comprehensive manpower

### data system.

Possible alternative systems were realistically circumscribed by such requirements as: (1) The information provided must be truly representative of a defined population (or populations); (2) The information collected must include at least the "Essential" Data elements required by policy makers; and (3) the information should be obtained and analyzed in the most efficient means possible, and preferably at a cost less than that of the National Register. The cost criterion also dictated that any proposed new systems be based on existing ones.

# Manpower Characteristics System

These principles led to the identification and examination of several alternative approaches in terms of practicality of collection, coverage, costs, flexibility, and so forth. After consideration of these alternatives, the following one was selected as the appropriate approach. This approach now called the Manpower Characteristics System consists of three separate and specific mechanisms or systems: A Comprehensive Doctorate Roster, based on the existing National Research Council efforts; a continuous periodic collection from a National Sample of Scientists and Engineers based on the 1970 Decennial Census; and Surveys of Nondoctorate Entrants to Science and Engineering, based on the present surveys of undergraduates conducted by the American Council on Education. This combination of collections will permit detailed periodic coverage of doctorates, as well as relatively adequate coverage of the total science and engineering population and the critical new entrants. Though each of the subsystems consists of periodic collections of data from individual scientists and engineers, it will not be possible to match or integrate specific individual scientists and engineers across the subsystems, because of the confidential nature of the records. However, it will be possible to integrate and analyze data on separate cohorts or populations.

The three separate parts of the Manpower Characteristics System are presently in varying stages of development as follows:

Comprehensive Doctorate Roster. The Office of Scientific Personnel of the National Research Council is now proceeding under contract with the National Science Foundation to develop a Comprehensive Doctorate Roster. Over many years, the Council with the support of the Foundation, and several other Federal agencies, has maintained a Doctorate Record File. This file consist of information collected from each person receiving an earned doctorate, including data on personal characteristics, educational background, and career plans. One of the first stages of the development of the Roster will be to collate this file of information with the records on doctorates contained in the several biennial (1954 to 1970) registerations of the National Register. Data on persons earning a doctorate from other than U.S. institutions and presently employed

in the United States will also be assembled. Using these sources, a universe of doctorates will thus be established. Under this plan, the Foundation hopes to have by the fall of 1972 a file containing the names, activity status, and current address of as many of the estimated 190,000 doctoral scientists and engineers as can be reached. Once established, the Comprehensive Doctorate Roster will be maintained on a current basis with respect to the inflow of recipients of newly awarded doctorate degrees from U. S. universities (to foreign and U.S. citizens). In addition, efforts will continue on including persons (U.S. and foreign citizens) receiving doctorates from foreign institutions and employed in the United States.

The Roster will then be available for general and special purpose data programs as needed. Without anticipating all uses of this resource, the operation will include routinely an annual survey of newly awarded doctorates, taken about a year after the award and covering the nature of post-award activity (other pertinent data will be available from the Doctorate Record File); a biennial collection of Essential Data from a cross-section sample of the entire science and engineering doctorate population, as well as a biennial updating of mailing addresses of the doctorate population. Additional data elements will be included as the need arises.

National Sample of Scientists and Engineers. The 1970 Census of Population collected information on the education, occupation, industry of employment, and other characteristics of a 20-percent sample of the entire U.S. population, including approximately 125,000 scientists, 240,000 engineers, and about 2.7 million persons with four or more years of college. This information will provide benchmark data on the science and engineering population, as it existed in 1970.

The Foundation is presently supporting a Postcensal Survey of Scientific and Professional Manpower directed at a 100,000-case subsample, drawn from the Census 20-percent sample, of persons in science and engineering occupations and persons with four or more years of college. The project is being carried out by the Bureau of the Census. The actual collection of detailed information took place in spring and summer 1972 and obtained 1970 and 1972 information to provide a more comprehensive picture of the 1970 science and engineering work force. (Mr. Seltzer will describe this project in greater detail).

In addition, the Foundation has supported the Bureau of the Census in obtaining an additional sample of 50,000 cases from the Decennial Census. These additional cases were surveyed in early summer of 1972 to obtain new addresses and selected items of data. Information collected from the two samples selected from the 1970 Census (100,000 from the Postcensal Survey and the additional 50,000 cases) will be examined using a set of screening criteria (age, level of education, and employment) to establish an adjusted science and engineering population. This population will then be used as the foundation for a series of periodic biennial examinations of the <u>total</u> science and engineering work force. Additional benefits accruing from the larger size sample include greater flexibility in cross tabulative data and an improvement in reliability, especially for smaller cells.

Survey of New Additions of Nondoctorates. An adequate collection of information on the scientific and engineering workforce must include data on persons entering the work force for the first time--primarily new graduates. The collection of such information is particularly important, as a supplement to the 1970 population of scientists and engineers obtained from the Census. The proposed comprehensive Doctorate Roster will provide a system for obtaining needed information on newly-awarded doctorates. However, a substantial data gap will exist on new entrants at the bachelor's and master's degree levels. Only a minority of those receiving science baccalaureates directly enter science activities; however, a sizeable proportion of new engineering graduates enter engineering. Many new baccalaureates enter graduate training (though not always in science or engineering fields), some enter other careers (high school teaching, non-technical fields, and even technician jobs), and some (mainly women) do not enter the labor force. At the master's degree level similar patterns exist, Few surveys have been undertaken to obtain information on the activities of new graduates, particularly below the doctorate level.

A basic system exists for obtaining information on new baccalaureates--the American Council on Education's longitudinal survey of college freshmen. Each year since 1966, a cohort of new freshman has been established and a series of follow-up surveys conducted on their undergraduate careers.

In mid 1972, the American Council on Education was awarded a development contract to develop a system of obtaining information on baccalaureates in science and engineering. The surveys of nondoctorate entrants will be based upon the cohorts of freshman sureyed each year by the American Council on Education. The goal of this part of the system include a series of continuing alternate year follow-up surveys of these freshman after graduation to determine trends in the career and educational patterns of new baccalaureates. These surveys will be carried out through the 1970's as a series of overlapping annual collections, including, coverage of a new cohort every two or three years, and on the alternate years coverage of a new cohort every two or three years, and on the alternate years coverage of preceding classes at 2, 3, 4, etc., years after graduation, including after several years coverage of a class of which a substantial proportion had earned a master's degree. Experience gained after several of these surveys could lead to a variation in this pattern.

In FY 1972 and FY 1973, a considerable amount of effort will be expended on the

development of these systems. In later years, as the systems mature, it will be necessary to continually carry out systems research and evaluation involving reexamination of the techniques of collection, methodology of analysis, and identification of subtle biases. For example, it is already realized that the method of collecting information on new nondoctorate entrants was chosen because a system existed which could be adapted to these purposes within the resources available. However, some new entrants will be missed, including non-degree and foreign scientists and engineers. It is hoped that the systems research activity will develop means to correct these shortcomings and identify other subtle biases arising out of new circumstances. To these and similar ends, a limited program of research and development on these subjects is proposed. A series of small projects will be designed to be carried out annually beginning in FY 1974.

Thus, the three separate, but interrelated, parts of the Manpower Characteristics system are being established to provide a continuing mechanism for collecting and analyzing information on the science and engineering population of the United States. It has all the features of the Register described above--collection of personal and professional information from individual scientists and engineers, which can be processed on a single time and on a longitudinal basis. Furthermore, the information obtainable from the system will be more truly representative of defined populations, will include essential data elements, and can be provided with less resources than the Register.

1/ Office of Management and Budget, <u>Special</u> Analyses, Budget of the United States, Fiscal Year 1972, p. 89.