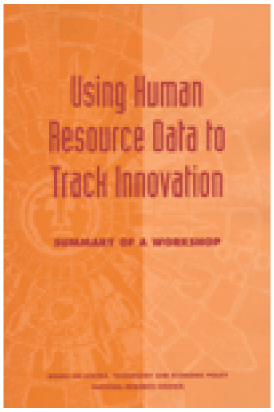

Using Human Resource Data to Track Innovation: Summary of a Workshop

Stephen A. Merrill and Michael McGeary, Editors,
National Research Council

ISBN: 0-309-50194-6, 80 pages, 6x9, (2002)

Using Human Resource Data to Track Innovation

SUMMARY OF A WORKSHOP

Edited by

STEPHEN A. MERRILL

and

MICHAEL MCGEARY

Board on Science, Technology, and Economic Policy

Policy and Global Affairs

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

NATIONAL ACADEMY PRESS • 2101 Constitution Avenue, N.W. • Washington, D.C. 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competencies and with regard for appropriate balance.

This study was supported by Contract No. NASW-99037, Task Order 103, between the National Academy of Sciences and the National Aeronautics and Space Administration, National Institutes of Health Contract No. 467-MZ-802221, and Department of Energy Grant No. DE-FG02-98ER30291. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-08424-5

Limited copies are available from:

Board on Science, Technology, and Economic Policy
National Research Council
500 Fifth Street N.W.
Washington, D.C. 20001
Phone: 202-334-2200
Fax: 202-334-1505

Additional copies of this report are available from National Academy Press, 2101 Constitution Avenue, N.W., Lockbox 285, Washington, D.C. 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

Adam Korobow
Program Officer

Sujai Shivakumar
Program Officer

Craig Schultz*
Research Associate

Camille Collett*
Program Associate

Chris Hayter
Program Associate

Michael McGeary*
Consultant

* Indicates staff for this project

Preface

The improved competitive performance of much of the U.S. industry in the 1990s derived from a combination of corporate strategies and public policies supportive of innovation, the latter including steady and conservative fiscal policy, economic deregulation, trade liberalization, relatively lenient antitrust enforcement, and the research investments of previous decades. These were the conclusions of an in-depth study of 11 manufacturing and service industries by the National Academies' Board on Science Technology and Economic Policy (STEP) published in 1999 (National Research Council, 1999a, 1999b).

Although cautiously optimistic about the future performance of the economy, the STEP Board articulated four concerns that have continued to guide much of its work: the availability of skilled human capital, the implications for research and innovation of some aspects of the extension of intellectual property rights, the adequacy of public and private investment in long-range research, especially in the physical sciences and engineering, and the adequacy of measures and statistical data to inform policy making.

The STEP Board's first effort to assess the utility and policy relevance of the government's data on innovation was a February 1997 workshop sponsored by the Sciences Resource Studies Division (SRS) of the National Science Foundation and summarized in *Industrial Research and Innovation Indicators* (National Research Council, 1997). In 2001, a committee formed by the STEP Board to study shifts in the allocation of federal research expenditures during the 1990s recommended several improvements in the collection, classification, and analysis of data on research and development (R&D) spending in both the private and public sectors (National Research Council, 2001). Another Academy panel recently reviewed the entire National Science Foundation portfolio of survey data on R&D and science and engineering personnel and recommended changes that would improve measurement of innovation (National Research Council, 2000b).

This volume is the summary of a second STEP workshop, chaired by board member Mark Myers, formerly chief technical officer of Xerox Corporation. The workshop explored how data on scientists, engineers, and other professionals—data on their training and skills, mobility and career paths, use of time, relationships across institutions and sectors, and productivity—can be used to illuminate aspects of innovation that current R&D, patent and other data, by themselves, do not fully capture.

In preparation for the meeting the STEP Board commissioned an exploratory paper by Paula Stephan, an economist at the Andrew Young School of Public Policy at Georgia State University. On November 23, 1999, the paper was presented to an audience of statisticians and economists, society and association representatives, government officials representing technical and statistical agencies and industrialists. Other presentations described applications of human resource data in research and the features of several federal government surveys containing human resource data. Participants also discussed ways to acquire and use new data and to link information from separate existing data sets.

The report does not present conclusions and recommendations of the STEP Board or of the Academies but does represent a faithful summary of the discussions of opportunities to improve understanding of industrial innovation and its outcomes through creative uses of information on professionals involved in the process.

The workshop was sponsored by the National Aeronautics and Space Administration, the Department of Energy, and the National Institutes of Health.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their contributions to the review:

Michael Finn, Oak Ridge Associated Universities
Bradford Jensen, U.S. Census Bureau

Carlos Kruytbosch, National Science Foundation (Retired)
David Roessner, Georgia Institute of Technology
Kenneth Troske, University of Missouri-Columbia

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the content of the report, nor did they see the final draft before its release. The review of this report was overseen by Robert McGuckin, The Conference Board, who was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Dale Jorgenson,
Chairman

Stephen A. Merrill,
Executive Director

Contents

Executive Summary

Data on scientists, engineers, and other professionals—their training, employment and mobility, structure of work and affiliations, and productivity—are an important but underutilized source of information on industrial innovation, a critical determinant of economic growth and productivity. Human resource (HR) data can supplement or compensate for the limitations of data on research and development expenditures, patents, and other innovation input and output indicators. They can also be used to evaluate government research and education programs. Ultimately, a better understanding of innovation characteristics and how they are changing can inform public policies to stimulate innovation and influence its direction, as well as to enable more people to benefit from it or to moderate its adverse effects on some groups in the population.

The principal official sources of data on professionals involved in the innovation process are surveys of the National Science Foundation (NSF), U.S. Bureau of the Census, and Bureau of Labor Statistics (BLS). Since 1993, several NSF survey data sets have been integrated into Scientists and Engineers Statistical Data (SESTAT), the most comprehensive and easily accessed source of information about the employment, education, and demographic characteristics of U.S.-resident scientists and engineers with at least a bachelor's degree. BLS and Census survey data are combined to produce the National Industry-Occupation Employment Matrix (NIOEM), which includes establishments in all sectors of the economy and all members of the scientific and technical labor force even below the baccalaureate level, although it does not provide demographic or educational attainment information. Under some circumstances data sets can be linked to enable a simultaneous examination of worker characteristics and firm characteristics. In addition, there are databases such as patent files and indexes to scientific publications that can be searched by individual. Some government research agencies, universities, and professional associations track the employment and work product of individuals affiliated with them; and, of course, scholars construct their own data sets for particular purposes.

SESTAT data show several trends in the deployment of PhD scientists and engineers in the United States—the increased role of industry as an employer relative to universities and government, the increased importance of the service sector relative to manufacturing industries, and the movement from the laboratory into non-R&D positions. These trends illustrate broader changes in the innovation process—the reduced role of government-financed R&D, the emergence of the service sector as a locus of much innovation, and the integration of R&D into other functions of the firm—strategy, management, and marketing. Perhaps the best examples of productive uses of HR data are research on the growth of biotechnology, which has no standard industry classification category, and research on alliances among firms and collaborations between firms and university researchers.

Much can be done to enhance the utility of human resource data. More precise information on employers and their locations would enable analysts to relate individual and firm characteristics and help compensate for the lack of business unit detail in the national R&D expenditure data. More information on what scientists and engineers do in firms would help illuminate the relationship of research to other functions. Information on scientists' and engineers' outputs (e.g., publications, conference proceedings, and patents) would help illuminate R&D spillovers between firms and across industries. And information on people in science and engineering (S&E) occupations who nevertheless lack a baccalaureate degree would help us understand a prevalent pattern in information technology.

Important advances in the analysis of innovation would come from linking human resource data with other data sets. For example, matching the name and location of respondents to the NSF surveys with Census establishment data could illuminate the relationship between firm size and innovation and between internal and external sources of innovation.

Finally, there is a need for new data on how people actually spend their time. Knowing what share of industrial scientists' time is spent in the laboratory, managing other employees' research, assessing the capabilities of other firms, collaborating with professionals outside the firm, or in production engineering would tell us a good deal about how innovation is taking place.

I

What Role for Human Resource Data in Tracking Innovation?

INTRODUCTION

The National Academies' Board on Science, Technology, and Economic Policy (STEP) conducted a workshop on November 23, 1999, to consider how more systematic exploitation of data on professionals—their training, mobility and career paths, functions in corporations, relationships across sectors, and productivity—could improve understanding of the process of innovation.

The premise of the workshop was that innovation—the invention, commercialization, and diffusion of new products, processes, and services—is an important determinant of economic growth, productivity, and welfare, and we want to understand better the factors that govern it. Ultimately, we want to use that understanding to adjust public policies to stimulate innovation and influence its direction to improve public health, enhance national security, and protect the environment as well as to foster economic development. Although it is a long way from a better understanding of the characteristics of innovation and how they are changing to policy prescriptions, measures informed by this understanding could take a variety of forms—increased public investment in research and training of scientists and engineers in particular fields, modifications of regulations that impede innovation, or adjustments to tax rates or intellectual property policies.

Innovation's role in improving the competitive performance of U.S. industry in the 1990s has been a principal focus of the STEP Board's attention, along with that of other economic analysts. Among other activities, STEP commissioned 11 industry case studies to determine the extent and sources of U.S. industrial resurgence from the mid-1980s to the latter half of the 1990s. Summarizing the findings regarding sectors as diverse as steel and semiconductors and food retailing and banking, David Mowery observed:

> The papers use an array of different measures to measure performance, and not all of them are calibrated against the performance of non-U.S. firms in these industries. Nevertheless, the overall portrait is one of stronger performance, not least in the ability of firms to develop and deploy new products and processes…. As many authors point out, firms have strengthened their ability to exploit their own or externally sourced innovations more effectively, rather than focusing exclusively or even primarily on improvements in their research or development capabilities (National Research Council, 1999b, pp. 3-4).

Ideally, the collection of data on innovation should be guided by a solid theoretical understanding of the process and its impact. Such understanding of causes and effects, while being developed, is not very far advanced. It is possible, however, to elaborate a conceptual framework that encompasses direct indicators of innovation, what we believe to be the principal influences on innovation, and its effects on economic performance as a way of cataloging innovation indicators and data. The broad categories of innovation information include

- influences on innovative activity, ranging from market conditions to investments in R&D and human resource capacity to the ways knowledge is communicated and used (inputs);
- innovation characteristics as evidenced by new products, processes, and services introduced in the market (outputs); and
- the effects of innovation on firms, workers, regions, and the economy as a whole (outcomes).

LIMITATIONS OF R&D DATA

The pattern, determinants, and effects of innovation involve variables that change over time, sometimes remarkably quickly and radically. At no time has that been more apparent than in recent years. At the workshop Paula Stephan, in introducing her commissioned paper, described several changes that are widely assumed to be occurring rapidly, to be substantial, and to have significant consequences for economic performance in the short term or in the longer run. These include (1) shifts in the industry and technology distribution of innovative activity, (2) shifts in the time horizon of innovative effort and investment, (3) changes in the organizational structure of innovative activity, and (4) changes in the location of innovative activity both within the United States and globally.

The presumed changes and some speculative concerns about their implications are as follows:

- *Distribution.* The sectoral and technological distribution of U.S. industrial research and innovation is shifting toward nonmanufacturing and new emerging industries and technologies. This is of concern to the extent that some industries are becoming less capable of attaining and sustaining long-term competititive advantage.
- *Orientation.* U.S. firms have been conducting less fundamental research with longer-term payoffs, focusing instead on more incremental innovative efforts with clear market applications and generally shorter time horizons. This may lead to a weakening of U.S.-located innovative capacity in the long term, although it may have enhanced firms' profitability in the near term.
- *Organization.* In-house R&D activities have been decentralized as firms have shifted away from the central R&D laboratory model. Decentralization of intrafirm innovative activity may have resulted in greater integration of R&D with corporate strategy and with other business units (product design, marketing, etc.), but unique technical resources of central corporate laboratories may have been dissipated in the process. Another organizational change is the increase in collaboration and outsourcing of research and innovation, not only among firms but also between firms and universities and government laboratories. This may improve the efficiency of innovative activities by reducing redundancies and accelerating implementation and diffusion, but it may entail a "hollowing out" of companies' innovative capacity, with possible negative effects on long-term competitiveness.
- *Location.* Firms in some industries have shifted R&D capacity abroad and entered into alliances with foreign firms, possibly gaining access to talent, technology, and markets but also somewhat reducing the likelihood of locating future activity in the United States. At the same time, foreign firms have invested more in U.S.-based R&D activities. As firms continue to outsource many of their innovation-related activities and work closely with universities and other public and private institutions, successful innovation appears to depend increasingly on geographically clustered networks of related organizations. Regions such as Silicon Valley, Research Triangle Park, and Austin, Texas prosper while regions lacking key infrastructure may experience economic stagnation.

These alleged changes in innovation patterns, even if real and substantial, are not permanent. Others will be identified in due course and become the focus of attention. Nevertheless, the changes described are the source of much of the current discussion of the need for improvements in industrial science and technology indicators and data. Consequently, they are useful partial tests of how well current innovation data serve our analytical purposes.

In part because they are the most consistently collected data and represented the best data time series related to innovation, R&D expenditures are often taken as the best or even surrogate indicator of innovation. In the conceptual scheme outlined above, R&D represent innovative effort and an important influence on whether new products and services and processes are developed and commercialized; but they are not a substitute for direct measures of innovation.

Paula Stephan and other workshop participants observed that the most commonly cited R&D data, from the National Science Foundation's Industrial Research and Development (RD-1) Survey, have other limitations as a source of information on innovation, and particularly on recent changes in innovation processes (see also Cooper and Merrill, 1998).

- *Distribution.* An effort in recent years to include nonmanufacturing firms in the survey has confirmed that there is substantial and growing R&D activity in service industries. But the survey reports expenditures only at the firm level, classified in 2- or 3-digit Standard Industrial Classification (SIC) categories, obscuring the composition of activity within large diversified firms, and precluding analysis of activity in emerging technologies or industries lacking an SIC category (e.g., biotechnology).
- *Orientation.* NSF surveys of both public and private R&D expenditures classify activities as basic research, applied research, and development, which are defined in standard ways. (The industrial survey no longer asks for a short-term/long-term breakdown of expenditures.) Corporate basic research did decline in the mid-1990s but only temporarily and partly as a function of growth in overall private R&D activity. Besides, it is not clear what bearing the basic/applied/development classification has on the concern that corporate R&D has shifted to shorter term, more incremental objectives.
- *Organization.* The collection of R&D spending data at the corporate rather than the business unit level is a severe handicap in assessing whether there has been a redistribution of activity and tighter integration of R&D and business strategy within most firms. Business R&D data show modest growth in contracting out and in contributions to universities in the 1990s (and the revenues have clearly become more important to the recipients), but there has been little increase in their share of all corporate R&D. Likewise, counts of consortia creation, R&D joint venture agreements, Cooperative Research and Development Agreements (CRADAs)

with federal laboratories, and mergers and acquisitions, while in some but not all cases increasing over time, fail to the indicate the origin, longevity, intensity and value of the collaborations. Moreover, these data are confined to formal, usually contractual relationships, while there is a dearth of statistical data on individual and informal transactions—technical communications between firms, customer-supplier exchanges, participation in scientific and technical meetings, and transactions involving transfers of knowledge by consultants, accountants, systems integration firms, and a variety of other sources.

- *Location.* The industrial R&D survey includes private and foreign-owned as well as publicly held companies but covers activities carried out in the United States only. Reporting of R&D funds spent abroad is voluntary. Moreover, again because the survey reports expenditures at the firm level geographic detail on the distribution of domestic expenditures is limited.

Although in the terms of the classification outlined at the beginning of this report, there are other kinds of innovation input or activity data—for example, counts of patents, counts of contractual collaborations, and counts of inventions or innovations—none goes very far in overcoming these limitations.

HUMAN RESOURCE DATA AND THE PROCESS OF INNOVATION

Data on professionals' skills, work, relationships, and productivity is an alternate or complementary source of information on industrial innovation. That is not merely because considerable data exist but because technological advance depends on human resources. In economic terms, technology is embodied in human as well as physical capital, and the interaction of scientists, engineers, and technicians is a principal means of technology diffusion. Changes in the distribution, orientation, organization, and location of innovative activity and capacity are reflected in changes in the training and mobility (education and career paths), the place and structure of work and affiliations (e.g., industry, occupation, allocation of time, consultancies, and informal collaborations), and the output and productivity of skilled personnel. Despite the obvious importance of human resources, national HR data have been underutilized and insufficently appreciated.

Discussion of uses of HR data at the workshop focused primarily on scientists and engineers with at least bachelor-degree-level training on whom there is fairly robust information from surveys (described in Chapter 2) sponsored by the National Science Foundation. But of course scientists and engineers are not the only contributors to innovation. It would be useful to consider what other kinds of professionals are instrumental to innovation and what information about them exists or could be acquired, but these topics were largely beyond the scope of the workshop.

CATEGORIES OF DATA

The workshop elaborated the following typology of professional characteristics primarily relevant to the science and engineering workforce that might have a bearing on innovation and on which survey data are or could be collected:

Training
- Educational qualifications and whether in or out of science and engineering
- In-service training or continuing education

Employment and Mobility
- Changing established firms or sectors
- Multiple institutions
- Domestic geographic movement
- International geographic movement

Structure of Work and Affiliations
- Allocation of time among functions (e.g., research, administration, information technology management, investment, analysis, etc.)
- Participation in interfirm or multifirm projects (consortia)
- Coauthorship of scientific or technical papers
- Consulting relationships
- Participation in other collaborations

Output/Productivity
- Patents applied for or received
- Publications
- Citations
- Salaries and other income (e.g., royalties, fees, etc.)

LINKS TO CHARACTERISTICS OF INNOVATION

Various workshop participants suggested ways in which these characteristics might be used to illuminate changes in innovation processes. The examples are illustrative rather than exhaustive.

- *Sectoral Distribution.* Paula Stephan pointed out that the non-manufacturing sector of the economy accounts for an even larger share of the industrial employment of scientists and engineers—about 45 percent—than its share of industrial R&D expenditures—about 25 percent. An examination of the roles of scientists and engineers in service sector firms would in all likelihood confirm that to a large extent innovative activity in that sector entails applications of information technology to customer service, marketing, inventory, and logistics rather than activities represented by standard R&D data. Similarly, the role of scientists, many of them academics, in forming and advising new firms has been critical to the emergence of industrial biotechnology and illustrates its close connections to publicly funded fundamental research. (See Chapter 3).
- *Time Horizon Orientation.* Subtle shifts in the focus of R&D and innovation within established firms might be illuminated by changes in the mix of personnel and their qualifications. For example, a reorientation toward short-term projects and incremental innovation might be associated with a change in the mix of scientists and engineers or of technical versus nontechnical professionals. Similarly, shrinking time horizons might be associated with technical professionals devoting less time to research and more to production engineering.
- *Organization.* The movement of technical personnel within firms could be a leading indicator of decentralization and the creation of teams of professionals with mixed qualifications and experience an indicator of the integration of R&D with other business functions—strategy, production, and marketing. The rise of outsourcing, alliances, and other collaborations is presumably associated with higher rates of copublication, coinvention, and consulting relationships. Another way to assess the incidence of collaborations and their importance to firms is to determine how senior technical personnel allocate their time between in-house activities and managing external relationships. Probing further, one could use data on individuals to shed light on the origins of collaborations and how they change over time.
- *Location.* Determining the distribution of scientists and engineers within and across firms and tracking changes over time would shed a good deal of light on the extent of regional clustering and globalization.

HUMAN RESOURCE DATA AND EFFECTS OF INNOVATION

Creative uses of human resource data could also explore links among characteristics of professionals working in firms, innovations, and traditional measures of performance of firms—sales, profitability, employment, and productivity. Stephan cited the work on industrial biotechnology of Lynne Zucker and Michael Darby, who show that the extent of collaboration between firm scientists and leading ("star") academic scientists is a powerful predictor of firms' success in terms of how many products they have on the market or in development, their sales, and their size and the value of their stock (Darby, 1999).

Donald Siegel of Nottingham University discussed a second important set of effects of innovation, on the workforce itself and how the workforce is managed. There is a strong theoretical case for believing that technological innovation is skill-biased; it increases the demand for highly educated and highly skilled workers because they have a comparative advantage in helping companies implement new technologies effectively (Siegel, 1999). Testing the hypothesis depends critically on data not only on specific technology adoptions but also on labor composition and relative compensation before and after the implementation of new technology. Siegel pointed out that a good deal of research has been on the manufacturing sector and the effects of process innovations, such as computer-aided design and manufacturing, robotics, flexible manufacturing systems, and just-in-time inventory systems. But much work remains to be done on the effects of service sector innovation and in distinguishing types of technologies with different workforce efforts, developing better measures of skill than educational level, and addressing intervening variables such as organizational and human resource management changes.

HUMAN RESOURCE DATA IN EVALUATING GOVERNMENT PERFORMANCE

Workshop participants also observed that data on professionals' career paths and productivity can supplement other means of evaluating government research and education programs. The link to government education grant, traineeship, fellowship, and research associateship programs is obvious. But federal research grants and contracts also indirectly support a significant share of graduate students in most fields of science and engineering. Government laboratories provide postdoctoral and training to scientists and engineers who leave to work elsewhere, often in industry. As a general matter, linking output and human resource data to government funding data provides an important avenue for measuring government performance. The Academies' Committee on Science, Engineering, and Public Policy (COSEPUP), in a series of workshops and reports, has discussed the importance of using human resources output as a performance measure for research under the Government Performance and Results Act of 1993 (National Academy of Sciences et al., 1999, 2001).

II

Principal Sources of Human Resource Data

The workshop identified four principal sources of data on professionals involved in the innovation process. First, the National Science Foundation, U.S. Bureau of the Census, and Bureau of Labor Statistics conduct national surveys yielding personal information on scientists and engineers or members of the workforce generally. Second, there are databases—for example, patent files and indexes to scientific publications—with relevant personal information that can be searched by individual. Third, some institutions track the employment histories and work product of individuals affiliated with them—for example, university graduates, professional association members, and government research agency grantees, trainees, and research associates. Finally, of course, scholars construct their own data sets for particular research purposes. The workshop was not intended to produce an inventory of these sources, but participants described the principal national surveys and cited examples of other sources.

NATIONAL SCIENCE FOUNDATION SURVEYS

Mary Golladay of the Science Resources Studies Division described the four NSF surveys and an integrated database that currently provide information on scientists and engineers educated or working in the United States.

Survey of Earned Doctorates (SED)

The SED is an annual census of all individuals earning a doctoral degree in the United States. The survey was begun in 1958 by NSF and cosponsored by the U.S. Department of Education, National Institutes of Health, National Endowment for the Humanities, and the U.S. Department of Agriculture. It includes information on demographic characteristics such as date of birth, marital status, education of parents, and geographic location of high school attended (Sanderson and Dugoni, 1999). There are also questions about field of training, sources of financial support during graduate education, and postgraduate employment plans. NSF issues an annual report which provides the same information just on the science and engineering doctorates (Hill, 1999).

Survey of Doctoral Recipients (SDR)

The SDR is a longitudinal demographic survey of science and engineering doctorate holders conducted biennially for the NSF and other federal agencies since 1973. In this survey, a sample of holders of doctorates in science and engineering earned at U.S. institutions is followed throughout their careers from year of degree award until age 76. Every 2 years, a sample of new S&E doctoral degree earners is added to the SDR from the SED. In 1999, for example, the sample frame included U.S.-earned S&E doctorates through the 1998 academic year. Detailed statistical tables in this report provide information on the number of scientists and engineers by demographic characteristic such as citizenship, place of birth, field of degree, and employment-related characteristic such as occupation, sector of employment, median salary, and various labor force statistics (e.g., unemployment rate).

Stephan, in her background paper, used data from the SDR to illustrate three trends in the deployment of skilled human resources that reflect changes in the structure of innovation. First, in all S&E fields there is a marked increase in the share of PhDs working in industry. Second, with the exception of chemistry, the share of PhDs employed in manufacturing industries has declined over time. Third, an increasing number of PhDs in industry are not engaged in R&D or R&D management.

National Survey of College Graduates (NSCG)

The NSCG was first administered in 1993[1] and biennially thereafter to a nationally representative sample of all college degree holders who were identified through the 1990 decennial census. The target population

[1]There was a hiatus in the late 1980s in the collection of data on S&Es when problems developed with the surveys based on the 1980 census (STPDS).

for this survey includes individuals in the United States as of April 1990 with a bachelor's degree or higher in any field, not just in science or engineering. In addition to including people with degrees earned at U.S. institutions, the NSCG also includes college degree holders who earned their degrees outside of the United States and were living in the United States in 1990. In 1993, those with science or engineering degrees and those without such degrees but working in S&E occupations were selected from the NSCG. These two populations are collectively referred to as the "S&E panel" of the NSCG. These same two groups were followed in the 1995, 1997, and 1999 rounds of the survey.

National Survey of Recent College Graduates (NSRCG)

The NSRCG has been administered biennially since 1974 to recent S&E bachelor's and master's degree recipients. The 1997 survey, for example, included those who earned bachelor's and master's degrees in science and engineering in the 1995 and 1996 academic years. Topics include educational experience before and after obtaining the sampled degree; graduate employment characteristics including occupation, salary, unemployment, underemployment, and post-degree work-related training; relationship between education and employment; and graduate background and demographic characteristics. The data may be used to understand the employment experiences of recent graduates such as the extent to which recent graduates entered the labor force, whether they were able to find employment, and the attributes of that employment. Results of this survey are presented separately for bachelor's and master's degree recipients and also separately for graduates of the two graduating class years.

Scientists and Engineers Statistical Data (SESTAT)

Since 1993 the SDR, NSRCG, and S&E panel of the NSCG have been integrated into SESTAT, the most comprehensive and easily accessed (http://srsstats.sbe.nsf.gov/) source of information about the employment, education, and demographic characteristics of scientists and engineers in the United States (Kannankutty et al., 1999). The SESTAT target population includes residents of the United States with at least a bachelor's degree and who, as of the reference date of the survey (i.e., April 15, 1993, April 15, 1995, April 15, 1997, etc.) were trained or working as a scientist or engineer, were less than 75 years old, and were not institutionalized. Not included in the sampling frames are individuals with associate's degrees in S&E fields or who are working in S&E occupations but lack bachelor's degrees. After 1993, the SESTAT surveys include only individuals whose degrees are from U.S. institutions and thus exclude immigrants with degrees from non-U.S. institutions who entered the United States after 1990. Some individuals have multiple chances of selection because they may have been included in the sampling frames for more than one component survey.

The principal variables in the SESTAT database are listed in Box 2-1. Access to some data is restricted to protect respondents' confidentiality.

BUREAU OF LABOR STATISTICS AND BUREAU OF THE CENSUS SURVEYS

Michael McElroy and James Spletzer, representing the Bureau of Labor Statistics, explained that the agency collects occupational employment statistics through three surveys, principally the Occupational Employment Statistics Survey, supplemented by the Current Population Survey (conducted jointly with the Census Bureau) and the Current Employment Survey, which are combined to create the National-Industry Occupation Employment Matrix.

Occupational Employment Statistics Survey (OES)

The OES program conducts a yearly mail survey of nonfarm establishments in order to produce employment and wage estimates for over 700 occupations. Data on self-employed persons are not collected and are not included in the estimates. The OES program produces these occupational estimates by geographic area and by industry. Estimates based on geographic areas are available at the national, state, and metropolitan area levels. The Bureau of Labor Statistics produces occupational employment and wage estimates for over 400 industry classifications at the national level. The industry classifications correspond to the two- and three-digit Standard Industrial Classification (SIC) industrial groups.

The OES program surveys approximately 400,000 establishments per year, taking three years to fully collect the sample of 1.2 million establishments. To reduce respondent burden, the collection is on a 3-year survey cycle that ensures that establishments employing fewer than 250 workers are surveyed at most once every 3 years. The estimates for occupations in nonfarm establishments are based on OES data collected for the reference months of October, November, or December.

The 1996 survey round was the first year in which the OES program began to collect wage rate data along with the occupational employment data in every state. In addition, the program's 3-year survey cycle was modified to collect data from all covered industries each year. Prior to

BOX 2-1
Examples of Variables in SESTAT

For the employed:
Primary job and salary
If previously retired
Type of employer: educational institution (by type); private for-profit; private not-for-profit; government (state/local or federal); self-employed
Supervisory responsibility, including number typically supervised directly and through subordinates
Relationship between work and highest degree, including reasons for employment outside the highest degree field
Typical work activities (in 14 categories), including primary and secondary work activities
Licensing and certification if required, recommended, or held
U.S. government support for research, including supporting agencies or departments
Second job, including occupation, salary, and relationship between work and highest degree field

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

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

Education

First bachelor's and two most recent degrees—level, degree field (major and minor), when awarded
Earlier education—date awarded high school diploma; associate degree(s)
Continuing education—post-degree college courses, reasons and field of study; employer financing

Other information

Family-related:
Marital status
Spouse's employment status; if working full/part-time, technical expertise required on job

continued

Children living at home (and ages)
Parents' educational attainment

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

Special modules

1993: Labor force status in 1988:
Type of employer and job
If different from current job, reasons for changing employer or job

1995 (SDR only): Post-doctoral experience:
Whether ever held a post-doctoral position
Number of post-docs held over career
Type of employer, including types of benefits offered
Whether current job was a post-doctoral position

1995 (NSCG and SDR only): Patent and publication activity:
Number of articles or other publications authored by respondent
Number of patent applications, patents awarded and commercializations attributed to respondent

1997: Alternative or temporary work experience:
Whether relationship to employer was alternative or temporary (consulting, contracting, etc.)
Reasons for such work arrangements
Whether benefits were provided, and if so, types of benefits

1996 the OES program collected only occupational employment data for selected industries in each year of the 3-year survey cycle.

Information contained in the survey is shown in Box 2-2.

Current Population and Employment Suveys (CPS and CES)

The CPS, a monthly survey of a probability sample of 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics, provides information on the employment and unemployment

Box 2-2
Occupational Employment Statistics Survey Data

National Occupational Employment and Wage Estimates
- Total employment by occupation
- Wages by occupation
- Occupational employment distribution by wage range

State Occupational Employment and Wage Estimates
- Total employment by occupation
- Wages by occupation

Metropolitan Area Occupational Employment and Wage Estimates
- Total employment by occupation
- Wages by occupation

experience of persons living in the United States. It is the primary source of information on the labor force characteristics of the U.S. population. Estimates from the CPS include employment, unemployment, earnings, hours of work, and other indicators. They are available by a variety of demographic characteristics including age, sex, race, marital status, educational attainment, occupation, and industry. CPS data are considered important indicators of the nation's economic situation and are used for planning and evaluating many government programs.

The CES, also a monthly survey, provides employment, hours, and earnings estimates based on payroll records of business establishments. The CES survey does not collect occupational information. Together, the CPS and CES fill in some of the gaps in coverage by the OES, such as self-, household, and farm employment.

National Industry-Occupation Employment Matrix (NIOEM)

The OES, CPS, and CES are combined to produce the National Industry-Occupation Employment Matrix as part of BLS's ongoing Occupational Employment Projections Program. The matrix shows occupational staffing patterns (occupation as a percent of the workforce) in 260 detailed industries and 513 detailed occupations. NIOEM includes establishments in all sectors of the economy and all members of the science and technology labor force of educational attainment, including those below the bachelor's level, in all academic disciplines (http://www.bls.gov/asp/

oep/nioem/empiohm.asp). It does not, however, have demographic or educational attainment information on individuals. Nor does it include those who have science and technology training but are in non-S&T jobs. For example, people with technical backgrounds who are top-level managers in industry or government are not captured in the "engineering, science, and computer systems managers" category, and other S&E-trained individuals are teachers, service personnel, writers, lawyers, etc. NIOEM categories do include groups not included in the SESTAT database—technicians and technologists and people in technical occupations where a bachelor's degree is not customarily required.

SESTAT and NIOEM each provide some information on the science and technology labor force and contribute to understanding the human resources required for science and technology in the United States (Kannankutty, 1999). NIOEM data give a broad view of the demand side of the technical labor market—jobs available as reported by establishments. SESTAT data give a more detailed picture of the supply of scientists and engineers with a bachelor's degree and above who are employed in the labor force. SESTAT shows that many people with S&E training have moved into the non-science and engineering labor force.

The two surveys are nevertheless not perfectly complementary. Although NIOEM includes employment data on technologists and technicians, complementary SESTAT data cannot be found for a large number of persons holding these jobs because they do not hold bachelor's degrees. The converse holds with regard to managers of the scientific and engineering enterprise: SESTAT can be used to identify scientists and engineers who are managers, but these people cannot be mapped into one specific category in the NIOEM (Kannankutty, 1999).

LINKING MICRODATA SETS WITH CONFIDENTIAL INFORMATION

J. Bradford Jensen, Director of the Center for Economic Studies (CES) of the Census Bureau, spoke about secure access sites for using Census data, the kinds of data available through the sites, and a project illustrating some of the research opportunities and constraints imposed by data confidentiality requirements. CES conducts empirical research on confidential microdata from the Census Bureau's regular survey and census programs.

Microdata are data at the level of the individual respondent, which might be a household, an individual within a household, an establishment, or a firm. Microdata can be used to create longitudinal or panel data that track respondents over time, which is impossible with aggregate data. Microdata also provide information about location and distance, which allows an assessment of spillover effects. Finally, microdata can be used to match individual data from other databases, for example, permitting the linkage of demographic (household and individual) data and economic (business establishment and firm) data.

Under Title XIII of the U.S. Code, the Bureau of the Census must keep the identity of respondents and the information they provide confidential. Data from demographic surveys are available in public-use files. It has proved impossible to create similar public-use files of data from economic surveys without either violating confidentiality requirements or editing the data to the point of rendering them unusable.

To enable researchers from outside the Census Bureau to use confidential microdata, CES has created a network of regional data centers, some with support from the National Science Foundation. The data centers, which are located in the Census Bureau's regional office in Boston, and at Carnegie-Mellon University in Pittsburgh, the University of California at Los Angeles, and the University of California-Berkeley, and Duke University, provide a secure site for researchers who obtain "Special Sworn Status" from the Census Bureau. Outside researchers may also have access at the CES in Suitland, Maryland.

To date, most Research Associates (as outside data users are called) have used the Longitudinal Research Database (LRD), which has longitudinal plant-level data from 1963 to the present, sometimes linked with related databases (e.g., Survey of Manufacturing Technology, Pollution Abatement Costs and Expenditures Database, Manufacturing Energy Consumption Survey, Industrial R&D). CES is broadening the LRD into a Longitudinal Business Database that includes economy-wide data, including the service sector, wholesale and retail trade, and finance, insurance, and real estate.

In the mid-1990s, CES created the Worker-Establishment Characteristic Database (WECD), in which worker characteristics and firm characteristics can be looked at simultaneously. WECD combines demographic information on workers from the long form of the 1990 Decennial Census with information on manufacturing plants where the workers were employed. A New Worker-Establishment Characteristic Database is being assembled that includes all industries.

Researchers also may now access confidential demographic microdata, avoiding the restricted geography and "topcoding" of income and other continuous variables of data in the public-use files. This could permit linking with the National Survey of College Graduates conducted by the Census Bureau for the NSF, as well as other surveys (e.g., Current Population Survey, Survey of Income and Program Participation, American Housing Survey, etc.).

Finally, it should be possible to link with outside data, because the identity of individual respondents is known. The Census databases include consistent data over a long period of time but cannot collect every question. Combining them with data in commercial databases or collected by researchers would increase the power of both data sets. For example, CES has teamed up with researchers at Carnegie Mellon to look at the impact of managed care on innovation in health care, which will link bibliometric and patent information with economic data on firms and hospitals. CES is also exploring with the American Medical Association (AMA) the possibility of linking AMA data on education and specialization of physicians with economic census information to study doctors' offices.

Jensen discussed one limitation on the Census economic data. Census does not survey very small establishments with less than 20 employees; it relies on administrative data from the Internal Revenue Service to capture some employment and payroll information on them. This reflects the primary focus of the Census Bureau on developing an accurate picture of aggregate economic activity as an input to the national product accounts of the Bureau of Economic Analysis. Very small establishments account for little economic activity. Nevertheless, this makes it harder to study business start-ups and to understand under what circumstances start-ups become large enough to join Census' sample frame. There are also problems in tracking mergers and acquisitions among small firms, because the amount of financial assets that changes hands may be trivial compared with the exchange of human capital, which is not measured. This is an issue that might be addressed by using human resources data to track the movement of innovative activity.

Julia Lane, Professor of Economics at American University, described the Longitudinal Employer Household Dynamics Project at the Census Bureau, a collaboration with John Abowd, Cornell University, and John Haltiwanger, University of Maryland. They are using administrative data from the Social Security Administration as the link record between information about individual persons, including earnings and employment histories, and economic data collected about their employers. With respect to scientists and engineers or highly educated individuals generally, because there are repeated observations on individuals, a relatively small initial sample frame becomes a much larger one. One could examine both cohort and temporal effects and career mobility.

III

Research Applications of Human Resource Data

Paula Stephan in her background paper used data from the Survey of Doctorate Recipients to document three trends in the deployment of scientists and engineers in the United States—the increased importance of industry as an employer of PhDs relative to universities and government; the increased importance of service industries as employers of scientific and engineering talent relative to the manufacturing sector; and the movement of PhDs from the laboratory into non-R&D positions. In addition, the workshop incorporated two sessions discussing examples of productive uses of human resources data in illuminating aspects of innovation—first, the growth of biotechnology and second the growth of alliances among firms and collaborations between firms and university researchers.

RESEARCH ON BIOTECHNOLOGY

The prominence of human resource data in studies of the emergence and growth of firms applying the scientific and technical advances of microbiology and biochemistry is no surprise. First, there is no standard industry classification for biotechnology and thus no ready-made universe of firms whose characteristics can be studied independent of their creators and managers. Second, those founders and managers tend to be scientists and engineers on whom there is considerable public information. Leading practitioners of this analysis include Lynne Zucker and Michael Darby, who have studied the emergence of biotechnology in the United States and abroad (*inter alia*, Darby et al., 1999; Zucker et al., 1994).

At the workshop other researchers described a series of studies that have examined, among other aspects of the development of industrial biotechnology, the following phenomena

- the contribution of basic research conducted in universities, government laboratories, and some large companies to the formation of new start-ups;
- tendencies not only to regional concentration of firms but also to regional specialization in certain types of biotechnology products or services;
- how scientists and engineers are employed within firms, not only in management and research but also in a variety of other functions such as quality assurance, regulatory compliance, and manufacturing design;
- how people with scientific and technical training populate service functions external to the biotechnology companies—venture capital, law, investment banking, and accounting firms;
- how biotechnology developments depend on informal networks of professionals that cut across public and private research and the nonprofit sector and that in some cases arise among graduate students and postdoctoral students before they become involved in entrepreneurial activity;
- the high degree of mobility and cross-fertilization among firms; and
- the feedback to university research via industry funding that tends to be associated with higher faculty productivity, albeit at some cost in openness of research.

Biotechnology in Maryland

Maryann Feldman of Johns Hopkins University presented her study of the genesis and evolution of the biotechnology industry in Maryland since the earliest firms were established in the early 1970s. Maryland is home to the third or fourth highest concentration of biotech firms in the United States. Feldman conducted her research by pulling together midrodata from a number of sources. She found 240 firms, 40 of them publicly traded. The median size is 14 employees; the mean size is just over 50 employees.

By finding out where the founders were employed before starting a firm, she discovered that most were spun off from large supplier firms, such as Litton Bionetics, Life Technologies, Inc., and Bethesda Research Labs, Inc., rather than universities. Those founded by academics were more likely to be from leading universities outside Maryland. The Walter Reed Army Institute of Research and the National Institutes of Health laboratories were also the source of founders of many firms, accounting for the fact that two of the core technologies in the Maryland biotech industry are vaccines and genomics/gene therapy.

Feldman had plans to repeat the study periodically. Because many of the companies tracked a year ago have gone out of business or merged, while other new firms have emerged, tracking the movement of scientific expertise will be one way to see what has been happening.

Biotechnology and the University of California

As director of the University of California's (UC's) industry/university cooperative research program, Suzanne Huttner oversees the university's biotechnology and other high-technology initiatives. She participated in a study of the role that basic research and research training at the university had played in the development of the California biotechnology economy. California has been the location of about one-third of U.S. biotechnology companies since the emergence of the industry; the UC system accounts for approximately 10 percent of National Institutes of Health extramural funding.

Led by Cherisa Yarkin, an economist at UC Berkeley, the study followed the people in the industry to find out where they were from, how they were deployed, and how they moved around. One-quarter of the firms in California were founded by a member of the UC faculty and 85 percent of the firms employed graduates of UC with advanced degrees in science and engineering. One hundred percent of those firms with 20 or more employees employed UC graduates. UC graduates with advanced degrees in molecular biology and related life sciences fields were working in most parts of a company including regulatory affairs, quality assurance, manufacturing, scale-up operations involving bioprocess engineering, and business development. Entirely new occupations are emerging that require expertise based on training in two or more disciplines. One such emerging field, bioinformatics, is experiencing severe labor shortages, because it requires advanced training in both molecular biology and computer sciences. The study found that advanced training in the life sciences is also a characteristic of those staffing investment banking, venture capital, law firms, and other services supporting the biotechnology industry.

Huttner concluded that failure to trace the education and employment backgrounds of workers in all sectors of the biotechnology industry, including the service and supply infrastructure of investment capital and legal services, as well as the production and business development workforce within firms, will result in badly underestimating the contributions of public investment, especially in research training, to the economy.

Scientific Networking and Entrepreneurial Success

Walter Powell of the Stanford University School of Education described his research on the role of scientific networks in the biotechnology industry (Powell et al., 1998). The project focused on how firms survive and grow by both doing research and absorbing ideas generated elsewhere. This line of inquiry, which involved assembling a detailed database on staffing, economic performance measures, and patterns of collaboration among firms, found high rates of mobility of scientists and complex patterns of interaction among them. Critical advances, such as the discovery of the BRACA-1 gene and development of the mouse model of Alzheimer's disease, involve collaborators from a variety of institutions—universities, government laboratories, nonprofit research institutes, research hospitals, biotechnology companies, and pharmaceutical houses. In the case of the BRACA-1 gene, the paper describing the discovery in *Nature* had 33 authors at 13 institutions; 11 of them changed employers in the 3-month interval between the submission of the paper and its publication.

Powell characterized the innovation process as competition among networks of scientists rather than competition between firms. Firms depend on recruiting researchers as employees or collaborators for access to these scientific networks, which speed the time in which they can bring products to market and generate revenues. Little is known about the formation and operation of these networks except that some of the most important relationships are formed in graduate school or postdoctoral appointments and many of them are too informal to be recorded as contractual ties.

Industry/University Relationships

Eric Campbell of the Massachusetts General Hospital and Harvard Medical School described research on the impact of industry funding on the research and other activities of a sample of basic biomedical and clinical faculty in the 50 universities that received the most funding from NIH in 1993. Approximately 28 percent of the life sciences faculty surveyed in these 50 most-research-intensive universities had industry support for

[1] This effect reversed, however, when faculty received more than two-thirds their research funding from industry. They published significantly fewer articles in the past 3 years.

their research. Faculty with industry support published significantly more articles in the most recent 3-year period,[1] compared with the average during their own careers and as well as compared with researchers without industry funding. Their teaching loads were about the same, and they devoted significantly more time to service activities. They were also more likely to have applied for, received, and licensed a patent and to have research become the basis of a start-up company.

Campbell estimated that studies that involve assembling a sample from scratch as quickly as possible take about 6 months of the time of the principal investigator and three or four graduate students and cost a minimum of $500,000.

RESEARCH ON COLLABORATIONS AND PARTNERSHIPS

Another productive area of innovation research using human resources data is also concerned with the transfer of technological information across institutional boundaries and ultimately into commercial activities, without being limited to a particular technology or industry.

University Technology Transfer

In the second part of his presentation, Donald Siegel summarized research on factors associated with the productivity of technology transfer offices of universities (Siegel et al., 1999). Commercialization of technologies by universities has increased greatly, from 300 patents in 1980 to 2,412 in 1997, from 276 licensing agreements in 1980 to 3,328 in 1997, and from 35 corporate start-ups in 1980 to 333 in 1997. Some universities do more than others. Siegel and his colleagues analyzed a survey of 113 university technology transfer offices affiliated with the Association of University Technology Managers, which has detailed information on sources and deployment of staff and their compensation together with information on university policies regarding the disposition of intellectual property and faculty and institutional involvement in start-up firms. They linked these data with data from NSF and the Bureau of the Census to create environmental controls. They found that adding environmental controls did not explain much of the variation in relative productivity among the offices, suggesting that individual characteristics and organizational and personnel policies matter more. Siegel noted that systematic assessment of the human resources characteristics and practices of successful university technology transfer offices could have practical policy consequences in identifying best practices, which in turn might facilitate more efficient spillovers of scientific knowledge.

Research Joint Ventures

Albert Link, an economist at the University of North Carolina at Greensboro, discussed recent research on research partnerships, both between companies and between companies and universities. He suggested that information on individual scientist and engineer participants is often revealing of the motivation and significance of the institutional linkages and the degree of collaboration. For example, firms infrequently assign top-level personnel to joint ventures, preferring to keep their principal human talent assets and core technology competencies proprietary. The characteristics of the researchers in interfirm collaborations can be a proxy for the intensity of firm involvement and a predictor of their economic consequences. Corporate collaborators with university researchers, on the other hand, tend to seek out star scientists with competencies that the firms are lacking. Indeed, it is somewhat misleading to speak of industry-university collaborations; the firms are generally much less interested in institutional than in individual capabilities.

Networks with Formal and Informal Elements

Diana Hicks of CHI Research spoke of possibilities of tying together databases on technical publications and patents, which are produced by people, with other human resources data to help measure the intangible elements of firm assets. For example, it appears that papers published by industrial researchers are more highly cited than those by academic researchers in certain fields of the biological and other sciences. If industry is the best place to do important research, who is doing it? What are the trends in the backgrounds of industrial researchers in those fields? An analysis of research publications in Great Britain shows that the growth is taking place at companies that publish one to 10 papers a year, indicating that employment of scientists with advanced degrees is increasing most rapidly in small companies, where their expertise can be more closely related to innovation (Hicks and Katz, 1997). Research activity in the service sector, as measured by publications, is even more concentrated in small firms. The percentage of papers coauthored with university researchers increased in nearly every industry apart from agriculture from the early 1980s to the early 1990s.

Turning to U.S. data, Hicks examined the science and technology linkages of DuPont along eight dimensions: co-authoring, co-patenting, patent-to-patent citation (each way), paper-to-paper citation (each way), and patent-to-paper citation (each way).[2] Such an analysis shows, for

---

[2]Paper-to-patent citations were not analyzed.

example, with whom DuPont scientists are publishing and upon whose research DuPont scientists are capitalizing. A similar analysis can be done of patents. The result is a visual map of the breadth and depth of the intellectual network in which DuPont operates—the institutions with which they collaborate in research and patenting and how often and the institutions whose work they have used or absorbed and vice versa, again with what frequency. The analysis shows how interlinked the structure of scientific research is.

Hicks cautioned, however, that the effort to pull the data together was very intensive and expensive, and a considerable investment would be needed to extend the analysis to more companies and include human resources data. Hicks estimated that it might take $3 million to $4 million to clean and integrate up to 10 years of data from the various citation and human resources databases and another $25,000 a year to maintain it, excluding the cost of getting data on sources of research funding. Others said that estimate might be low, perhaps very low, especially for annual operating costs. On the other hand, Stephan observed that there is a substantial unmeasured cost in failing to improve the data already collected and their linkages, because they are revealing less and less about trends in industrial innovation.

IV

Enhancing the Utility of Human Resource Data

In discussing directions for human resource data development, workshop participants pointed out that there are undoubtedly numerous productive uses of the current SESTAT, BLS, and Census data. For some purposes, however, existing indicators are inadequate and new data are needed. For example, despite the extensive data on the production of degrees earned in science and engineering fields at the bachelor's, master's, and doctoral levels, surprisingly little is known about what they do in their jobs every day and over their careers, especially if those careers are in industry. In innovative sectors such as information technology technical work is often performed by people without formal science or engineering educations or without bachelor's degrees. For other purposes, the data exist but are not being fully exploited. There are also opportunities to link data collected by public and private institutions in more productive ways. But there are limits including budget constraints and privacy protections that will shape choices among potential improvements in human resource data collection and analysis.

Charlotte Kuh of the National Research Council staff commented that these choices deserved more systematic thought in light of the considerable cost of assembling large data sets and the constraints entailed in protecting the confidentiality of individually identifiable information. What do we really want to know about the relationship of human resources to innovation? What are the important questions? What reasonably strong conceptual models need testing? How can data systems be made sufficiently flexible to address policy issues that may arise in the future? The suggestions offered by workshop participants are simply candidates for further evaluation. They do not reflect any coherent intellectual framework for priority setting.

EXPANDING CURRENTLY COLLECTED SURVEY DATA

Several workshop participants perceived a need to derive more information on subjects of current surveys, particularly scientists and engineers working in industry:

- More precise information on employers and their locations (as is available on academic scientists and engineers) would enable analysts to relate individual and firm characteristics. In particular, if SDR respondents employed in companies were asked to indicate the industrial classification of their establishment (plant, laboratory, etc.) this would help overcome the lack of business unit level R&D expenditure data.[1]
- More information on what scientists and engineers do in firms would help illuminate the relationship of research to other functions—strategy, finance, production, and marketing—highly relevant to successful innovation. (See below for a suggestion for obtaining even more detailed information on activities.)
- Information on scientists' and engineers' outputs and public activity (publications, conference presentations, involvement in consortia or other collaborations, etc.) would help illuminate R&D spillovers among firms, between industries, and across sectors. The 1995 NSCG and SDR contained a module on patenting and publishing (Morgan et al., 2001). Such a module, modified to reflect changing patterns of publishing, patenting, and collaboration, might be included periodically in the SESTAT surveys.
- Data on stock options, which are prevalent in high technology industries, would fill a growing gap in the information on professionals' nonwage compensation.

Some workshop participants maintained that it is desirable to expand the NSF definition of the S&E workforce and the information obtained about certain categories of scientists and engineers, although the costs of such steps would have to be considered. In particular,

[1]An alternative to including a long, unwieldly SIC code listing with the survey questionnaire is to ask the respondents the name of the sub-unit of the national organization and then conducting a post-survey coding of the answers into fine SIC codes. This has not been implemented because of resource limitations.

- Very little information is available about the work of scientists and engineers who occupy managerial positions that nonetheless use knowledge and skills from post-secondary scientific and technical training to direct or influence innovation. Examples are plant manager, division executive, strategic planner, patent attorney, and business developer.
- People in S&E positions who nevertheless lack an S&E baccalaureate degree are not included in the SESTAT surveys. This is presumed to be prevalent in information technology fields (e.g., computer programming and network administration); but a workshop participant familiar with many biotechnology start-up companies said that it is also the case in that industry that a significant share of the technical workforce lacks a BA or BS degree.
- Finally, it was observed that the NSF SESTAT data on the web and on compact disc should be constructed to permit longitudinal analysis, which is not currently the case.

FACILITATING LINKAGES BETWEEN DATA SETS

Workshop participants generally agreed that more important advances in analysis of innovation would come from linking human resource data with other data sets. Paula Stephan cited the example of matching the name and location of respondents to the SDR and SED with Census establishment data providing detailed information on characteristics of the firms. Such a link could illuminate the relationship between firm size and innovation as well as the internal versus external sources of innovation by small firms. The lack of information from HR data alone about the resources available to individual scientists and engineers could be overcome by linking the SDR data with firm R&D expenditures or with databases such as the CRISP file of National Institutes of Health research grants. Other promising linkages are between SDR data and publication data in the International Scientific Index and between SDR and patent data.

In the judgment of a number of researchers at the workshop, the creation of the regional Census Bureau centers at which qualified investigators can access confidential microdata has been crucial to efforts to link economic datasets and their expansion should be considered. Their utility would be greater, however, if they provided access to OES and other Bureau of Labor Statistics data sets.

CREATING NEW DATA

Jim Adams observed that one of the most important but least well understood questions about innovation processes is how people actually spend their time. What percentage of industrial scientists' time is spent conducting research or development, managing others' R&D research, assessing the R&D and technological capabilities of other firms (e.g. acquisition on collaboration candidates interacting with customers), collaborating and communicating with professionals outside the firm, or in production engineering? How much time is devoted to continuing education or simply keeping up with the research field? Time-use surveys, in which individuals are interviewed over a period of time or asked to maintain diaries, are an accepted way of addressing such questions and have been used in a variety of economic contexts such as to measure unpaid work as input for satellite accounts to national economic accounts and to help evaluate income and welfare policies (National Research Council, 2000a). This research method has not been used to better understand the innovation process, however. Some form of time-use survey is a candidate for a SESTAT special module.

A final suggestion was that federal agencies consider sponsoring additional targeted surveys of key professional groups of interest—for example, biotechnologists—collecting information on activities, output, relationships, and compensation well beyond that solicited in the NSF surveys of scientists and engineers.

References

Cooper, Ronald S., and Stephen A. Merrill, eds. 1998. "An Assessment of National Data of U.S. Industrial Innovation." Unpublished paper, National Science Foundation Division of Science Resources Studies.

Darby, Michael R., Qiao Liu, and Lynne G. Zucker. 1999. "Stakes and Stars: The Effect of Intellectual Human Capital on the Level and Variability of High-Tech Firms' Market Values." National Bureau of Economic Research Working Paper No. 7201.

Hicks, Diana, and Sylvan Katz. 1997. "The Changing Shape of British Industrial Research." STEP Special Report No. 6.

Hill, Susan T. 1999. *Science and Engineering Doctorate Awards: 1998*. NSF 00-304. Arlington, VA: National Science Foundation.

Kannankutty, Nirmala. 1999. *SESTAT and NIOEM: Two Federal Databases Provide Complementary Information on the Science and Technology Labor Force*. NSF 99-349. Arlington, VA: National Science Foundation.

Kannankutty, Nirmala, and R. Keith Wilkinson. 1999. *SESTAT: A Tool for Studying Scientists and Engineers in the United States*. NSF 99-337. Arlington, VA: National Science Foundation.

Morgan, Robert, Carlos Kruytbusch, and Nirmala Kannankutty. 2001. "Patenting and Invention Activity of U.S. Scientists and Engineers in the Academic Sector: Comparisons with Industry," *Journal of Technology Transfer*, vol. 26, Jan. 2001, pp. 173-83.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. 1999. *Evaluating Federal Research Programs.* Washington, D.C.: National Academy Press; 2001. *Implementing the Government Performance and Results Act for Research.* Washington, D.C.: National Academy Press.

National Research Council. 1997. *Industrial Research and Innovation Indicators*. Washington, D.C.: National Academy Press

National Research Council. 1999a. *Securing America's Industrial Strength.* Washington, D.C.: National Academy Press; 1999b. *U.S. Industry in 2000: Studies in Competitive Performance* Washington, D.C.: National Academy Press.

National Research Council. 2000a. *Time-Use Measurement and Research: Report of a Workshop*. Washington, D.C.: National Academy Press.

National Research Council. 2000b. *Measuring the Science and Engineering Enterprise, Priorities for the Division of Science Resource Studies*. Washington, D.C.: National Academy Press.

National Research Council. 2001. *Trends in Federal Support of Research and Graduate Education*. Washington, D.C.: National Academy Press.

Powell, Walter W., Kenneth W. Koput, Laurel Smith Doerr, and Jason Owen Smith. 1998. "Network Position and Firm Performance: Organizational Returns to Collaboration in the Technology Industry." Unpublished paper.

Sanderson, Allen R., and Bernard Dugoni. 1999. *Summary Report 1997: Doctorate Recipients from United States Universities*. Chicago, IL: National Opinion Research Center, University of Chicago.

Siegel, Donald. 1999. *Skill-Based Technological Change: Evidence from a Firm-Level Survey*. Kalamazoo, MI: W.E. Upjohn Institute Press.

Siegel, Donald, David Waldman, and Albert Link. 1999. "Assessing the Impact of Organizational Practices on the Productivity of University Technology Transfer Offices: An Exploratory Study." NBER Working Paper No. 7256. July.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1994. "Intellectual Capital and the Birth of U.S. Biotechnology Enterprises." NBER Working Paper No. 4653.

APPENDIXES

Appendix A

Workshop Program

8:30 **Welcome and Introduction**
Mark Myers, Xerox Corporation and Board on Science, Technology and Economic Policy

8:45
Session I: **Framing the Issues and Objectives**
Paula Stephan, Georgia State University
Discussant: *James Adams, University of Florida*

9:30
Session II: **Research on Biotechnology**
Walter Schaeffer, NIH, Chair
Susanne Huttner, University of California
Maryann Feldman, Johns Hopkins University
Walter (Woody) Powell, Stanford University
Eric Campbell, Harvard Medical School and Massachusetts General Hospital

11:00
Session III: **Research on Collaborations and Partnerships**
Kathie Olsen, NASA, Chair
Donald Siegel, Arizona State University West
Al Link, University of North Carolina, Greensboro
Diana Hicks, CHI Research

Sample questions for Sessions II and III:

What innovation analysis using human resource data has been done and what issues has it illuminated?
What have been the sources of data and difficulties encountered?
What linkages among data sets have been possible and productive?
What opportunities are there for other uses?
What difficulties do you anticipate?

12:30 Lunch

1:15
Session IV: **Opportunities and Obstacles to New Data Uses and Coordination**

Nancy Kirkendall, Energy Information Agency, Chair
Brad Jensen, Center for Economic Studies, Bureau of the Census
Mary Golladay, National Science Foundation
Michael McElroy, Bureau of Labor Statistics
Julia Lane, Bureau of the Census, American University, and the Urban Institute

2:30
Breakouts **What's Possible? Where Do We Go From Here?**

Sample questions for Breakout Groups:

To assess differences in the role of human capital in innovation, should the analysis and techniques applied to biotechnology be applied to other industries or technologies? Which?
What HR data have been collected but not exploited?
What linkages can be made between HR and other data sets? How? Can this be done without disrupting time series or losing other valuable information?
Should federal agencies or supported institutions track personnel (laboratory employees, PIs, trainees, graduate assistants, research fellows and associates, etc.)?
What aspects of program evaluation would benefit from use of HR data?

4:30
Session V: **Reporting and Summation**

Appendix B

List of Participants

James Adams
University of Florida

Clifford Adelman
Department of Education

Martin Apple
CSSP

Tom Arrison
National Academies

Eleanor L. Babco
Commission on Professionals in Science and Technology

Anita Balachandra
U.S. Department of Commerce

Grant Black
Georgia State University

Jennifer Bond
National Science Foundation

Mark Boroush
National Institutes of Health

Bob Burkart
Industrial Research Institute

Eric Campbell
Partners Healthcare System

Roger Chalkley
Vanderbilt University Medical School

Dennis Chamot
National Academies

Mike Champness
NAB

Melvyn Ciment
Potomac Institute for Policy Studies

McAlister Clabaugh
National Academies

Ron Cooper
U.S. Small Business Administration

Marc Cummings
U.S. Department of Commerce

Roman Czujko
American Institute of Physics

Stephen Dahms
San Diego State University

Walter Diewald
National Academies

Nancy Donovan
U.S. General Accounting Office

Sidney Draggan
U.S. Environmental Protection Agency

Chad Evans
Council on Competitiveness

Maryann Feldman
Johns Hopkins University

Pam Flattau
Flattau Associates, LLC

Brian Flinchum
University of North Carolina, Greensboro

Barbara Fraumeni
Bureau of Economic Analysis

Fred Gault
Statistics Canada

Mary Golladay
National Science Foundation

Cary Gravatt
Department of Commerce

Jong-on Hahm
National Academies

Diane Hicks
CHI Research

John Horrigan
National Academies

Diana Hoyt
NASA

Susanne Huttner
University of California

John Jankowski
National Science Foundation

Brad Jensen
Bureau of the Census

Elmima Johnson
National Science Foundation

Gretchen Jordan
Sandia National Laboratories

Robin Kawazoe
National Institutes of Health

Maryellen Kelley
NIST

Nancy Kirkendall
Energy Information Agency

Sylvia Kraemer
NASA

Charlotte Kuh
National Academies

Julia Lane
American University

Al Link
University of North Carolina, Greensboro

Bill Long
Business Performance Research Associates

Tom Mahoney
West Virginia Manufacturing Extension Partnership

Brian Mannix
The Manufacturers Alliance/ MAPI Inc.

Michael McElroy
Bureau of Labor Statistics

Michael McGeary
McGeary & Smith

Stephen A. Merrill
National Academies

Jarle Moen
Harvard University

Percy Moleke
Human Sciences Research Council, South Africa

Youhyun Moon
Korean Embassy

Robert Morgan
National Science Foundation

Mark Myers
Xerox Corporation

Robin Nazzaro
U.S. General Accounting Office

Judy Nyquist
National Academies

Kathie Olsen
NASA

Inja Paik
Department of Energy

Linda Parker
National Science Foundation

Benjamin Powell
University of Pennsylvania

Walter (Woody) Powell
Stanford University

Phanish Puranam
University of Pennsylvania

Stephen Quigley
Science Policy & Management Consultant

Samuel Rankin, III
American Mathematical Society

Alan Rapoport
National Science Foundation

Scott Rayder
House Committee on Science

Diane Raynes
U.S. General Accounting Office

Proctor Reid
National Academy of Engineering

Robert Rich
American Chemical Society

Daniel Rodriquez
U.S. General Accounting Office

Sally Rood
Federal Laboratory Consortium

Tom Rozzell
National Academies

Rosalie Ruegg
NIST

Walter Schaffer
National Institutes of Health

Craig Schultz
National Academies

James Schuttinga
National Institutes of Health

Scott Shane
University of Maryland

Donald Siegel
School of Management

Emily Smail
National Academies

Jim Spletzer
Bureau of Labor Statistics

Paula Stephan
Georgia State University

Debbie Stine
National Academies

Miron Straf
National Academies

Peter Syverson
Council of Graduate Schools

Barbara Torrey
National Academies

Wm. J. Valdez
U.S. Department of Energy

Debbie Van Opstal
Council on Competitiveness

Bruno van Pottelsberghe
Free University of Brussels

Philip Webre
Congressional Budget Office

Scott Weidman
National Academies

Douglas P. Wilson
Eli Lilly and Company

Appendix C

Using Human Resource Data to Illuminate Innovation and Research Utilization

PAULA E. STEPHAN

Andrew Young School of Policy Studies
Georgia State University

Issues paper prepared for a workshop of the
Board on Science, Technology, and Economic Policy
National Research Council

Grant Black provided valuable research assistance on this project.

SECTION I. INTRODUCTION

Substantial evidence exists that widespread changes are occurring in patterns of innovation. One consequence of these changes is that traditional measures, such as patent counts and research and development expenditure data, are increasingly unable to illuminate R&D activity in the United States. "Without substantial change in the content and coverage of data collection, our portrait of innovative activity in the U.S. economy is likely to become less and less accurate." (Mowery, 1999, p. 46). A key element of this change, although not the only element, is the increased incidence of collaboration both of a formal and informal nature that is occurring across institutions. By their very nature collaborative arrangements blur the boundaries between organizations and make it difficult to relate inputs, such as R&D expenditures of a firm, to outputs, such as patent counts.

This paper explores the use of human resource (HR) data concerning scientists and engineers to illuminate innovation and research utilization. For policy purposes it is important to gain an understanding of the extent to which this can be done, not only because of the aforementioned failure of current measures to capture changes occurring in the system but also as a way of gaining better insight into the social return to investments in science and engineering. For example, the federal government invests billions of dollars annually in funded research. Because grants are often awarded to individual principal investigators, human resource data and links of human resource data to outcomes, such as patents, make it possible to ascertain the eventual outcome of the investment and consequently make some inference concerning its economic impact. Similarly, federal labs train a significant number of young investigators who eventually leave to work elsewhere. The subsequent performance of these trainees could provide one measure of the contribution that federal labs make to overall performance.

The plan of this paper is as follows. Section II summarizes changing patterns of research and development and comments on gaps in our ability to measure innovative activity. Section III defines what is meant by human resource data and describes the data that are readily available. Examples of what can be learned from the use of human resource data to illuminate changes in innovation as well as provide insight in areas concerning the innovative process where substantial gaps exist are drawn by using the Survey of Doctorate Recipients (SDR). The section concludes by examining what could be learned if the HR data that are available were to be linked with other databases. Section IV looks at lessons learned from studies of biotech firms. Section V examines citation analysis.

In this discussion we are particularly interested in (1) what we can learn from data concerning the deployment of human resources as well as what we can learn from indicators such as publications that have individuals as a fundamental unit of analysis; (2) what can be learned from studying collaborative and sequential[1] relationships that often are identified by using human resource data; and (3) what we could learn if we had better human resource data or the ability to link together two or more existing databases, such as the SDR with firm data.

SECTION II. CHANGING PATTERNS OF RESEARCH AND DEVELOPMENT

At least four broad changes have occurred in the structure and organization of innovative activity in the United States: (1) a decreased role for federal funding of R&D; (2) a change in the industrial distribution of innovative activities; (3) a shift of resources toward development activities and away from basic research; and (4) a change in the organization of research.

[1] Here we use the term sequential to connote either the source of an innovation or the impact the innovation has on subsequent innovation.

Change in the Public/Private Mix of R&D

Industry has been the largest source of R&D funding in the United States for nearly two decades and its share continues to grow, approaching two-thirds of all R&D expenditures by 1998. This trend reflects both the increased growth of industrial R&D and a decrease in federal spending on R&D in terms of constant dollars in the late 1980s to the early 1990s. The decrease at the federal level since the mid-1980s has been largely a result of declining expenditures on R&D for defense. Federal civilian expenditures for R&D, on the other hand, which were relatively stable in the early 1980s in constant dollars, have increased in the late 1990s. This is largely a result of increased spending on health-related R&D, primarily through the National Institutes of Health.

Changes in the Industrial Mix of R&D

The most striking change in the industrial mix of innovation in the United States is the increased role played by the service sector. In the early 1980s R&D performance by nonmanufacturing industries made up less than 5 percent of total industry R&D performance. But beginning in the early 1980s this began to change dramatically. By 1991 the service sector accounted for nearly 20 percent of all the R&D performed in the manufacturing and service sector of the U.S. economy. Since then the service share has declined a bit but is still more than triple what it was only 15 years ago. The situation is, however, more dramatic than these statistics indicate since in "many nonmanufacturing industries that are essential to the development and diffusion of information technology, R&D investment is difficult to distinguish from operating, marketing, or materials expense" (Mowery, 1999, p. 46). This is reflected by the fact that although nonmanufacturing firms account for about 25 percent of the industrial R&D expenditures, they employ something like 45 percent of scientists, engineers, and S&E managers in industry.

A Shift Toward Development

Another trend observed in R&D data is a shift away from research toward development. "The upturn in real R&D spending that has resulted from more rapid growth in industry-funded R&D investment is almost entirely attributable to increased spending by U.S. industry on development, rather than research." (Mowery, 1999, p. 45). By 1997, eight out of every $10 spent by industry on R&D were directed toward development (National Science Board, 1998, Tables 4-6 and 1-17).

Changes in the Organization of Research

The organization of industrial research in the U.S. has also undergone substantial change. Four trends characterize the change: (1) increased reliance on external R&D, such as that performed by universities, consortia, and government laboratories (Mowery, 1999, p. 44); (2) increased collaboration in the development of new products and processes with domestic and foreign competitors and customers (Mowery, 1999, p. 44); (3) a decentralization of in-house R&D activities (Merrill and Cooper, 1999); and (4) the movement of innovative activities to functions in the firm typically not thought of as being drivers of innovation. The latter is fueled in part by the development of technologies that, as noted above, impact the operation and marketing of the firm's production. Because these changes contribute to the growing inadequacy of traditional measures to describe innovative activity, we examine each in some detail.

The trend toward increased reliance on external R&D is undeniable.[2] Firms outsource R&D to other firms or, in a more aggressive mode, acquire R&D through acquisitions.[3] Cisco Systems is an oft-cited example of the latter, acquiring start-up companies as an R&D strategy. One consequence of this strategy is that the role of scientists in the firm is not to perform R&D but to assess the R&D capabilities of possible acquisitions. Again, standard R&D data fail to classify such individuals as engaged in R&D and thus undercount R&D activity in the United States.

Since the passage of the National Cooperative Research Act (NCRA) in 1984, nearly 600 formal research joint ventures (RJVs) have been filed with the U.S. Attorney General and the Federal Trade Commission (Leyden and Link, 1999, p. 575). A not insignificant number of these RJVs include federal laboratories as partners. Link and Leyden estimate that overall the rate is 8.7 percent (p. 577). As a result of the growth of RJVs traditional R&D data portray innovative activity less accurately today than in the past. The fact that associated firms typically allocate a portion of a scientist's time to the collaborative project suggests that our knowledge of innovative activities resulting from RJVs could be enhanced by focusing on human resource deployment.

U.S. companies also enter into numerous collaborations with foreign companies. Most of these international alliances are with Western European companies although alliances between U.S. and Japanese companies are also

[2] See Badaracco, 1991; Hagedoorn, 1993; Hamel, 1991; Saxenian, 1994.

[3] Friar and Horwitch (1986, p. 77) studied 10 leading U.S. R&D companies. Although none was planning to extend its internal R&D activities, five intended to increase the acquisition of technology acquired through licensing, by forming joint ventures, or by fully acquiring firms with the needed technology resources.

widespread (Mowery, 1999, p. 44). While domestic consortia of firms focus their efforts most consistently on research, many of the international alliances focus on joint development, manufacture, or marketing of products.

Firms also outsource to universities. In recent years the proportion of university research that is funded directly by industry has grown from approximately 2 percent in 1960 to over 7 percent by 1997 (Mowery, 1999, p. 45; National Science Board, 1998, Table 4-6).[4] Trends in outsourcing to universities are clouded, however, by the fact that although industry funds an increasing share of university research, contributions to universities are not a significantly growing share of industry-funded R&D (Merrill and Cooper, 1999).

Universities have been not only a source of innovative ideas for established firms but also midwives to new firms formed by faculty members joining with venture capitalists. Nowhere has this been more evident than in the area of biotechnology,[5] although other examples, for instance software and lasers, exist. University founders and researchers often have their cake and eat it too, maintaining their university jobs while they work in industry. In other instances they move back and forth between universities and firms, taking sabbaticals at companies (Powell and Owen-Smith, 1998, p. 263). Numerous federal programs also exist promoting cooperation between universities and industry. For example, the National Science Foundation (NSF) has programs that promote university-industry collaboration, and in some instances funding requires that NSF-supported centers have an industrial component (Powell et al., 1998, p. 256).

Changes in the law have also encouraged federal labs to develop alliances with industry and develop arrangements whereby R&D activity can be outsourced to private labs. Federal labs also join research consortia that involve firms in the private sector. The Clinton administration's 1993 "defense conversion initiative" opened up formerly off-limits defense-related research to commercialization (Powell et al., 1998, p. 256).

Organizational changes have also occurred within the firm as some firms have shifted away from the central R&D lab model choosing not only to outsource research but in many instances to locate research activities at the plant level. This adds to the fuzziness of current R&D data since

[4] Universities are not only receiving a larger and larger proportion of their research funds from industry. University faculty increasingly are seeking patent protection for research performed within the university. One contributor to this is the eightfold growth in less than two decades in the number of university technology transfer offices.

[5] Audretsch and Stephan (1999) find that 50 of the 101 scientific founders of the 52 biotech firms they study were in academe at the time they founded the firm. At the time the firm went public, 35 of these founders remained working full time in academe.

the location of where the actual innovation is developed less and less corresponds to corporate headquarters. Moreover, the growth of mergers and acquisitions makes it increasingly difficult to associate R&D activity with firm output. This is because the survey instrument that collects R&D data is fielded to the firm rather than the business unit. This results in attributing all of the firm's R&D spending to the firm's industry classification. Thus, if 51 percent of the firm's business is in computer sales and 49 percent is in computer services, all of the R&D expenditures are attributed to the former category. Human resource data may help solve the "location" problem because they usually contain the address of the individual. HR data could overcome the line of business problem if the industry code of the plant or other establishment were made part of their record.

A final organizational change occurring within firms is the movement of innovative activities to functions in the firm not typically regarded as drivers of innovation. One example, given above, is the assignment of scientific personnel to evaluate and seek R&D through mergers and acquisitions. Another example is the involvement of technically teamed personnel in marketing and distribution. The important innovations that firms make in these areas are generally missed in standard measures of R&D. HR data could provide insight into these innovative activities by examining the deployment of S&E-trained individuals in non-R&D jobs.[6]

More generally, the improved competitive performance of many U.S. industries has come not only from the development of new technologies but also from "the more effective adoption and deployment of innovations" (Mowery, 1999, p. 42). These capacities are not measured by traditional R&D indicators such as patent counts and expenditures. They include "investments in human resources and training, the hiring of consultants or specialized providers of technology-intensive services, and the reorganization of business processes." (Mowery, 1999, p. 46). Creative uses of human resource data could illuminate industries that have high technology absorption capacities and could aid in our understanding of the strong performance enjoyed by a number of industrial sectors in recent years.

Powell and Owen-Smith (1998, p. 266) argue that some of the structural changes enumerated above have occurred because knowledge is increasingly located in networks of relationships and access to such networks is a key to competitive survival. In other work Powell and coauthors have shown that firm value is positively and significantly related

[6] Of course, it does not necessarily follow that individuals trained as scientists and engineers working outside of R&D in industry are using their training on the job. They may have accepted these jobs in the absence of employment opportunities in research.

to network access. Traditional R&D indicators fail to measure these networks as well as the access that individual companies have to networks.

These structural changes mean that traditional indicators of R&D as well as the traditional unit of analysis, the firm, are becoming less relevant to the study of innovation. But that is not the whole story. It is not just that the formal organization of R&D is changing. Ample evidence exists that knowledge spillovers play an important role in innovation and that traditional measures fail to capture the effects of these spillovers.[7] These data inadequacies are becoming apparent at a time when we tout the economic growth enjoyed by the United States as being "knowledge-based." We are attributing growth to inputs that, because of organizational change occurring within firms and the development of outsourcing and collaborative ventures, are increasingly difficult to measure accurately. In short, the changes outlined above result in a blurring of boundaries and a blurring of roles. Measures of innovation designed when firms were discrete firms and universities were strictly universities fail to portray these changes adequately.[8]

Even without the changes noted above the traditional measures of innovative activity, namely patent counts and R&D expenditures, reveal little to investors and analysts concerning the knowledge base of firms. As Lev (1999) notes, firms report nothing on a regular basis other than their R&D expenditures. This makes it difficult to evaluate companies, particularly companies that are knowledge-based. There is no way, for example, of determining the closeness of the science link or to evaluate the quality of the link.

SECTION III. USE OF HUMAN RESOURCE DATA

Human Resource Data: Definition and Availability

Broadly speaking, human resource data refer to data collected on individuals either working in or trained in the field of science and technology. Although such data can and are collected on a case-by-case basis as well as by professional societies and universities, six primary sources for HR data in S&E exist in the United States.[9] These are briefly summarized

---

[7] Spillovers are often examined by studying the relationship between a measure of innovative activity of the firm and the research expenditures of universities and other organizations in close geographic proximity. The rationale for expecting them to be bounded is that tacit knowledge is difficult to communicate in writing but is facilitated through face-to-face communication. See, for example, Jaffe, 1989; Acs et al., 1992.

[8] Powell and Owen-Smith (1998, p. 266) do a good job of summarizing these changing boundaries.

here along with the target population that each addresses. We then use one of them, the SDR, to illuminate several of the trends discussed above and as a way of exploring how, with certain linkages and additions, the data could be used to illuminate other trends existing in patterns of innovation.

The NSF directs considerable resources towards gathering information on the scientific and engineering workforce who have completed their PhD training in the United States. These data are collected in two complementary ways. First, the Survey of Earned Doctorates is administered to all individuals receiving a doctoral degree in the United States, regardless of field. This survey, begun in 1958, is administered by the awarding university and forms the basis of a census of all individuals who received their doctoral training in the United States. The census, referred to as the Doctoral Records File (DRF), was begun early in the twentieth century and was originally constructed from administrative records. Since 1957 survey data have been available on field of training, financial support during graduate education, employ-ment plans, and an array of demographic characteristics including date of birth, marital status, education of parents and geographic location of high school. The SDR is a biennial survey of a sample of individuals whose records are contained in the DRF and who indicated at the time that they received their doctoral degree that they intended on staying in the United States.[10] The intent is that the data be longitudinal and that individuals remain in the frame until the age of 75. The data capture individuals trained as scientists and engineers who are working outside their field of training as well as, until quite recently, individuals with doctoral degrees outside S&E. Thus the linguist who received a PhD in English but is now working in an information tech-nology field was included in the survey until financial considerations recently led National Endowment for the Humanities (NEH) to dis-continue their support for sampling the humanities.

Individuals who have not received training at the doctoral level clearly contribute to innovation. This is particularly the case in the areas of engi-

[9] A University of California study has tracked the career paths of PhDs in biochemistry, computer science, electrical engineering, mathematics, political science, and English over a ten-year period. The study group is composed of all of those receiving a PhD between July 1982 and June 1985. The results of the study for mathematicians and biochemists were reported by Nerad and Cerny (1999).

[10] Many more scientists and engineers indicate that they have plans to work abroad than actually do. In an exceptionally creative use of data linkages, Michael Finn and coauthors (1995) matched the SED records to Social Security records to estimate the number of individuals who say they have plans to leave but do not actually leave or who subsequently return. They used a similar procedure to examine whether individuals who said they definitely planned to stay in the United States actually stayed.

neering and computer information and technology. Human resource data are collected on nondoctoral-trained individuals (as well as doctoral-trained individuals) through two additional surveys: the National Survey of College Graduates (NSCG) and the National Survey of Recent College Graduates (NSRCG). Since 1993, both surveys have become biennial. The sampling frame for the NSCG is drawn from all college-educated individuals in the most recent decennial census regardless of occupation reported in the census. Follow-up biennial surveys include college-educated individuals trained and/or working in science and engineering. Thus, for example, the Russian physicist who immigrates to the United States but cannot find a job in S&E is included in the NSCG as is the physicist who works on Wall Street.[11] The sample also includes the linguist who works in information technology. By using the census for the basis of the sampling frame, the methodology includes individuals working in the United States who received their training outside the country.[12]

The NSRCG provides informa-tion about individuals who recently obtained bachelor's or master's degrees in S&E. The population surveyed includes all individuals under the age of 76 who received bachelor's or master's degrees in an S&E field within a 2-year period prior to the survey reference date from a U.S. institution. In addition to information concerning education and employ-ment status, the survey collects data on such variables as primary work activity, occupation, and salary. Information from these three surveys (SDR, NSCG, and NSRCG) has been integrated into the SESTAT database, available on the web or on CD-ROM. The SESTAT database allows for analyses of different components of the S&E population.

A question that readily arises when using these or any other HR data sources is, "Who constitutes the S&E workforce?" If analysis is restricted to individuals trained in S&E, the linguist who makes the transition to an S&T occupation is missed but the individual trained in physics working on Wall Street is included. If the analysis is restricted to those working in S&E, the linguist is included but the physicist missed. There appears to be no ready answer to this question of definition, but users are cautioned

[11] The predecessor to the NSCG was the 1982 Post Censal Survey. The sampling frame for this survey was different from the 1993 NSCG, however, being drawn from individuals identified as being in scientific and engineering occupations in the 1980 Census. The 1993 NSCG sample, by contrast, was drawn from all college-educated individuals regardless of occupation reported in the 1990 census.

[12] Given that the sampling frame is based on college education, the NSCG also includes individuals who received their doctoral training in the United States but left, only to return. It also includes medical doctors who, unless they receive a joint MD-PhD, are excluded from the SDR.

to be aware of underlying definitions when using the data. For example, the NSCG could be and has been used to study individuals trained out of the field working in information and technology as well as the proportion of the highly trained financial community who received their training in science. Furthermore, to date none of the NSF surveys tracks individuals who are technically trained but do not have a baccalaureate degree.

Information on those working in S&E occupations can also be obtained from two databases collected by the Bureau of Labor Statistics (BLS). The best known of these databases is the Current Population Survey (CPS), a monthly survey of approximately 50,000 households. Sampled units are asked for basic demographic information concerning all persons residing at the address and detailed labor force information for all persons 15 or over.[13] Included are questions related to level of education as well as detailed occupational codes. From the CPS one can obtain information on individuals working in S&E occupations by level of training. One cannot, however, identify individuals trained in S&E.

The Bureau of Labor Statistics also collects data concerning employment by occupation from establishments. Known as the Occupational Employment Statistics (OES) program, these data are collected yearly on wage and salary workers in nonfarm establishments to produce employment and wage estimates for over 750 occupations. The OES program surveys approximately 400,000 establishments per year, taking 3 years to collect the entire sample of 1.2 million establishments. Data are released at the aggregated level.

Broadly speaking, from the six sources described above we are able to obtain information on the training and deployment of individuals working in S&E occupations as well as individuals trained in S&E occupations. Where the surveys are longitudinal, we are able to observe changes over time and thus can examine mobility and earnings patterns over the life cycle. As a general rule, these sources contain little if any information on output measures other than salary or on patterns of collaboration with others working in the field of science and engineering.

[13] To improve the reliability of estimates of month-to-month and year-to-year change, 8 panels are used to rotate the sample each month. A sample unit is interviewed for 4 consecutive months, and, after an 8-month rest period, for the same 4 months a year later. Each month a new panel of addresses, or one-eighth of the total sample, is introduced.

The Survey of Doctorate Recipients

Several of the trends discussed in Section II are apparent from an examination of human resource data. For illustrative purposes we use data from the SDR, that span the period 1973 to 1993.[14] We use these data to illuminate three trends: (1) the increased importance of industry as an employer of PhDs in the United States; (2) within industry, the decreased importance of the manufacturing sector as an employer; and (3) and the increased deployment of industrial PhDs in non-R&D (or R&D management) positions.

Table C-1 presents the data by field and by year in summary form. The specific categories of interest to us for this study are (a) the percent in industry; (b) the percent in manufacturing of those in industry; (c) the percent in services and "other" of those who report jobs in industry; and (d) the percent in R&D or R&D management of those with positions in industry.[15] Because deployment varies considerably by field of training, we present the data for six fields.[16] The trends noted above hold in almost all instances across fields. In many instances the change is most noticeable in the early 1980s. Irrespective of field, we find an increase in the percent of PhDs working in industry. In math by 1993 21.7 percent of those trained in the field reported holding a position in industry, almost a three-fold increase over the 20-year period. In engineering and chemistry approximately one out of two PhDs worked in industry by 1993 and the proportion approached this in computer science. This increase in the deployment in industry has come largely at the expense of employment at PhD-granting institutions and reflects in part the poor job market conditions in the academic sector over much of the period.

---

[14] These data were available at Georgia State University under a licensing agreement at the time this paper was written. The 1995, 1997, and 1999 SDR data are available at the national level and could be incorporated into this analysis.

[15] Changes in survey design and execution affect the types of comparisons that can be performed and the interpretation of observed differences. For example, beginning in 1991 the sample size of the SDR was reduced due to funding constraints and more effort was invested in follow-up. This resulted in a reduction in non-response compared to the 1980s. These modifications and the low-response rate, particularly during the 1980s, compromise the robustness of comparisons that can be drawn over time. At the same time, the SDR is the primary source of national-level employment information for PhDs educated and working in the United States.

[16] The data are restricted to include those with good responses located in the United States. We exclude individuals reporting military employment and those who report that they are retired or out of the labor force.

TABLE C-1 Summary of Doctorate Recipients Data by Field and Year Characteristics of Employed Scientists and Engineers

Year	% in Industry	% in Manufacturing of Those in Industry	% in Service and "Other" of Those in Industry	% in R&D and R&D Management of Those in Industry
		Life Science		
1973	11.3	74.5	23.5	75.8
1979	13.7	61.9	34.7	59.6
1983	17.0	54.3	41.3	57.8
1989	21.0	53.9	39.2	61.3
1993	24.4	—	—	53.3[b]
		Chemistry		
1973	39.4	90.9	6.6	85.8
1979	43.9	81.8	14.2	80.0
1983	46.5	72.6	23.0	75.8
1989	48.8	77.2	19.9	75.8
1993	49.7	—	—	62.9[b]
		Physics and Astronomy		
1973	23.4	69.9	17.2	90.5
1979	28.6	65.2	27.5	78.3
1983	34.8	49.7	31.3	77.2
1989	34.4	48.8	38.3	76.5
1993	30.7	—	—	53.3[b]
		Math		
1973	7.9	59.9	21.6	80.1
1979	14.2	45.0	36.0	66.5
1983	19.4	39.8	39.9	67.9
1989	20.9	35.3	46.2	59.4
1993	21.7	—	—	35.2[b]
		Computer Science		
1973	[a]	[a]	[a]	[a]
1979	[a]	[a]	[a]	[a]
1983	45.1	49.0	45.3	83.1
1989	43.6	42.3	46.0	75.0
1993	44.8	—	—	38.9
		Engineering		
1973	44.0	73.1	17.6	83.8
1979	48.9	65.9	25.1	74.7
1983	52.2	53.7	37.8	70.3
1989	51.3	53.7	35.2	70.4
1993	51.8	—	—	56.3[b]

[a] Unweighted cell size below 30.
[b] The 1993 data are not comparable to the earlier years because counts include only those engaged in R&D; R&D management as a work activity was not distinguished from a generic management category in 1993.
SOURCE: Survey of Doctorate Recipients, NSF, 1973–1993.

Consistent with the patterns noted above, we see that among those working in industry the percent in manufacturing has declined over the period of observation.[17] For example, in physics and astronomy two out of three PhDs in industry were working in manufacturing in 1973; by 1989 that share was about one-half. The trend is even stronger in math, engineering, and computer science. Even in chemistry, where the vast majority of PhDs working in industry were in manufacturing in the 1970s, we see a decline, although the decline had begun to reverse itself by 1989.

The industrial sectors where the increase has been dramatic include service and "other," combined here because of the ambiguity of definitions used across the years.[18] In all but the life sciences, the propor-tion in industry working in these combined sectors more than doubled during the 16 year interval portrayed in Table C-1. In the life sciences the percent grew by more than 60 percent.

The final trend shown in Table C-1 is the decline in deployment, among those working in industry, in R&D and R&D management activity. This pattern is consistent with the observation that innovative activity is increasingly moving out of the lab and into other positions within the firm. Unfortunately, the 1993 survey failed to collect data on R&D management and thus the data for 1993 are not comparable with that for 1973-1989.

R&D activity, as measured by expenditures, is heavily concentrated in a small number of states (National Science Board, 1998, pp. 4-30). For example, one-half of the $177 billion spent on R&D in the United States in 1995 was expended in just six states: California, Michigan, New York, Massachusetts, New Jersey, and Texas (National Science Board, 1998, pp. 4-30). The top 11 states (adding Illinois, Pennsylvania, Maryland, Ohio, and Washington) perform two-thirds of all R&D. By contrast, the bottom 20 states produce less than 5 percent of the R&D conducted nationwide in 1995.

---

[17] Unfortunately, the 1993 survey did not collect information on industrial classification.

[18] The industrial classification used by the SDR changed substantially between 1983 and 1989. In particular, in 1973 and 1979 three-digit SIC industry codes were used. In 1983 four-digit codes were used. These were simplified considerably in the 1989 questionnaire where only two-digit codes were used. The tremendous increase in the proportion of PhDs reporting that they worked in the service sector between 1983 and 1989 and the decrease of those reporting working in the "other" sector suggests that when the classification was simplified, a considerably larger proportion of PhDs classified themselves as in the service sector. The sectors supposedly excluded from "other" across all periods are construction, manufacturing, mining, transportation, communication and utilities, wholesale and retail trade, and finance, insurance, and real estate. In 1991 the industrial classification was no longer done by coders but instead by the respondent. In 1993 the survey collected no information on industrial classification directly although the name of the employing institution is part of the record.

TABLE C-2 Geographic Distribution of Doctorate Scientists and Engineers in the Labor Force

	All Fields				
	Percentage of Labor Force in Selected Years				
Top Eleven States	1973	1979	1983	1989	1993
California	11.08	12.05	12.61	13.67	13.82
Illinois	4.34	4.39	4.44	4.22	4.07
Massachusetts	4.29	4.24	4.19	4.76	4.63
Maryland	5.46	5.73	5.48	5.92	6.23
Michigan	3.43	3.22	3.17	3.26	2.91
New Jersey	4.69	4.66	4.91	4.72	4.84
New York	9.67	9.10	8.50	7.85	7.27
North Carolina	—	—	—	—	2.81
Ohio	4.47	4.07	4.20	4.05	3.63
Pennsylvania	5.42	4.74	4.81	4.81	4.64
Texas	5.15	5.59	5.74	5.69	6.11
Virginia	2.60	2.92	2.98	2.94	—

SOURCE: Survey of Doctorate Recipients, NSF, 1973–1993.

The deployment of PhDs is also heavily concentrated, although not to the extent that R&D expenditures are. In 1993, for example, the top six states employed 43 percent of all PhDs; the top 11 employed 60.7 percent and the bottom 20 states 8.6 percent. Little change in the concentration of PhDs' employment, at least as measured by the SDR, has occurred over the 20-year period 1973–1993. In 1973 the top 6 states employed 41.3 percent, the top 11 employed 60.7 percent, and the bottom 20 employed 8.5 percent. These figures, however, mask certain changes that have occurred in deployment among the top states. Most notably, as we see from Table C-2, the gap between the top employing state, California, and the second state, New York, has widened considerably during the 20-year interval.

These data provide one means of examining geographic deployment—focusing on the U.S. doctoral-trained labor force—but fail to reveal the deployment patterns of foreign-trained PhDs. More importantly, they tell us nothing about employment patterns of the nondoctoral S&E workforce. Many of the changes noted above, however, undoubtedly reflect changes in the deployment of this portion of the workforce and we could undoubtedly learn something about these changes if these data were readily available by geographic location.

The blurring of boundaries between industry and academe and the extent of knowledge spillovers from academe to industry and vice versa make

TABLE C-3 Percentage of Doctorate Scientists and Engineers Employed in Industry Who are Located in the Same State as their PhD Institution

All Fields				
1973	1979	1983	1989	1993
All Doctorates				
25.6	25.0	25.9	26.1	26.2
5–6 Year Cohort				
25.2	28.3	27.2	27.7	29.3

SOURCE: Survey of Doctorate Recipients, NSF, 1973–1993.

the geographical deployment of scientists and engineers of particular interest. This is especially the case in situations where tacit knowledge plays a key role. One way in which knowledge transfers are fostered between industry and academe is through the placement of graduate students. Not only do new PhDs bring new ideas, but they help to build and maintain effective networks between industry and academe.

This raises the question of whether data on PhDs can shed light on changing patterns of deployment of new PhDs. We are particularly interested in changes in the percent of PhDs that accept industrial em-ployment within the same state of training.[19] Table C-3 summarizes these data. We see that while there has been no change for all doctorates when analyzed jointly, there is a slight increase for the 5-6 year out cohort, suggesting that the proclivity of newer PhDs for working for industry in their state of education has increased.

Linking HR Data with Other Data

As revealing as these data are concerning changes in innovation practices, they leave numerous unanswered questions that might be answered if available HR data were linked to other databases. Consider what is known versus what could be learned regarding scientists and engineers working in industry. Usually only one question is asked concerning industrial classification and this question is not consistently ascertained over time nor is it coded consistently over time. Furthermore, no infor-mation is available on characteristics of the employer. In contrast, we know much more about characteristics of employers for PhDs working in academic in-

[19] This is obviously an imprecise measure given that many cities spill across state lines.

stitutions. Much of this information (PhD-granting, Carnegie classification, etc.) comes not from the respondent but instead from matching the code of the employing academic institution with data collected on educational institutions. A similar matching could occur for industry employment. For example, respondents in both the SED and SDR surveys are asked to supply the name and location of the employing oganization. By matching this information with Census establishment data we could get more detailed information on firm characteristics including size.[20] Information provided by such a link could illuminate the relationship between firm size and innovation as well as the source (external vs. internal) of innovation in small firms. Such data could also go a long way to addressing the "line of business" issue mentioned earlier. [21]

Only one measure of output has been consistently collected in human resources databases of scientists and engineers and that is salary. We have little information on non-salary components of income, including stock options. More importantly, we have no indication of the respondents' productivity, as measured either by article counts and the citations associated with these articles or by patent counts.[22] Neither do we have information on the productivity of the firm for which the in-dividual works, as measured by traditional indicators of firm per-formance. Information on such dimensions could potentially be obtained by linking HR data with other databases. For example, Levin and Stephan (1991) had the SDR linked with publication data from the Institute for Scientific Information (ISI). Similar linkages could be made concerning patents.

Not only do we know very little about output measures; we also know little about inputs. The individual-based data that are collected provide minimal information about resources that are available to scientists and engineers, for example in the form of research grants from funding agencies or research budgets of firms. In theory, such information could be obtained by linking HR data to databases such as the Computer Retrieval of Information on Scientific Projects (CRISP) file of the National Institutes of Health which provides information on grants.

Another deficiency is that the existing HR data do not allow us to examine the degree to which scientists wear "multiple" hats in the sense that they work for more than one institution—an increasingly common

[20] Firm size has been collected in the SDR since 1995.

[21] Respondents are asked to supply the actual location where they work, which is often different from that of the corporate headquarters. Employer name is not coded for respondents working in industry but is retained in the record.

[22] The exception is that the 1995 SDR asked a question with regard to the number of papers and articles authored since 1990. It also asked a question with regard to whether the respondent had been named as an inventor on any patent application since 1990.

phenomenon given the amount of outsourcing from industry to academe and the number of academics who work with start-up firms.[23] Perhaps more importantly, we are unable to determine from survey data with whom scientists collaborate and thus cannot use the data to study characteristics of collaborative patterns such as geographic proximity and fields of complementarity. Nor can we determine how individuals are recruited into collaborative relationships, especially those in industry. Data linkage might enable us to trace how investments made by government labs in training researchers spill over to other sectors as the trainees leave to take positions outside the government. Data on patterns such as these would not only give us insight into changing patterns of innovation. They would also help us evaluate the impact of government programs designed to train scientists and foster the productivity of existing scientists and engineers.

SECTION IV. LESSONS LEARNED FROM THE STUDY OF BIOTECHNOLOGY

By far the best example of what can be learned by examining linkages based on human resource data comes from the study of biotechnology firms. Zucker and Darby have contributed a great deal to our under-standing of what can be learned through the careful analysis and linkage of data. Stephan has also examined several issues that can only be studied by linking individual with firm level data. Here we summarize the approaches and results of each.

Zucker and Darby have constructed a rich database and used it to advance our understanding of how and why new firms in biotechnology are established and locate in certain areas. Their work also informs our understanding of how firms and scientists outside these firms benefit from collaboration. The construction of this database has several elements. A key component was the construction of a measure of intellectual capital in biotechnology. This was done by identifying leading researchers, termed "stars," on the basis of the number of genetic sequence discoveries reported up to 1990 for which they were an author. Characteristics of these scientists, such as employing institution, were determined as well as characteristics of their coauthors regardless of whether the latter qualified to be stars (Zucker et al., 1994). Zucker and Darby have also collected considerable data on biotechnology firms and on research resources located in the same geographic area of the firm. Included in the latter data are the proximity of highly rated university departments and the level of spending on R&D.

[23] Since 1993, however, the SDR has asked for information on a secondary employer.

The construction of this type of database has contributed to a number of insights into the development of the biotechnology industry and the human resource dimensions of this development. Key findings include, but are not limited to, the following:

- Over time the proportion of stars and active collaborators working primarily in universities rather than firms has declined significantly, from nearly 100 percent initially to less than 50 percent by 1989.
- The growth and geographic location of intellectual capital was the principal determinant of the growth and location of the biotechnology industry (Zucker et al., 1994, p. 29).
- The extent of collaboration by a firm's scientists with stars is a powerful predictor of firm success as measured by products in development and on the market as well as the number of employees (Zucker and Darby, 1995b).
- Commercial involvement by stars is associated with increased research productivity as measured by article citations (Zucker and Darby, 1995b, p.18).
- The higher the quality of the star, the shorter is the time that the star remains at a university before moving into the biotechnology industry, other things being equal (Zucker et al., 1997).

Stephan's research focused on biotechnology firms that made an initial public offering during the hot market of the early 1990s. She used the prospectuses of the 50-odd firms to ascertain the names of the in-dividuals with scientific training affiliated with the firm. Although especially interested in scientists who give an academic address as their primary employer, she also collected data on full-time employees of the firm who are listed in key positions as well as the names of founders of the firm. Stephan then determined the citation counts of the scientists and key demographic information such as date of birth, country of birth, educational training, work history, geographic location of primary employer, and whether or not a Nobel Prize recipient. Various databases were linked and, in some instances, the scientists were asked for missing pieces of information. Stephan used the CRISP data to measure stock prices over time and the Investnet CDA database to determine the extent to which "insiders" engage in trades and profit taking.

Key findings from this work, which could only be ascertained through a human resources lens, include the following:

- Although a substantial number of university-based scientists participate in networks that are geographically bounded, approximately 70 percent of the links between biotechnology companies and university-based scientists are nonlocal (Audretsch and Stephan, 1996).

- Reputation of the scientists is positively related to day-one performance of the initial public offering. Proceeds raised also relate positively to reputation (Stephan, 1999).
- Scientific founders who come from academe are older and more highly cited than those who come from drug companies (Audretsch and Stephan, 1999).
- Approximately 10 percent of the university-based scientists held sufficient amounts of stock to qualify as "insiders" by the Securities and Exchange Commission. Among this group, it was not uncommon to engage in stock market activity that yielded handsome capital gains (Stephan and Everhart, 1998).

The work of Zucker and Darby and of Stephan shows the richness of results that can be obtained by linking data on scientists to indicators of their productivity such as citations and then linking this information either directly or by geographic indicator to firm data. No other industry appears to have garnered such attention and for no other industry have such intricate linkages based on human resource data been constructed.[24]

SECTION V. CITATION ANALYSIS

Publications

On a much larger scale, the bibliometric work of Hicks and her coauthors provides insights into changing patterns in innovation and, although the data are not based on human resource survey data, scientists and engineers generate the data for this work on authorship patterns. Hicks and coauthor Katz, through the use of bibliometrics, identify several changes occurring in the production of scientific papers. Key to the methodology

[24] Sleeper (1998) examines characteristics of founders of de novo firms in the laser industry. She finds that one-quarter of the founders came from university and government labs; close to half from the laser industry; a sixth from industry outside of lasers and an eighth are not identified. De novo firms have a higher exit rate than established firms that went into laser production. Within the de novo class of firms, those with founders from the laser industry have the highest survival rate followed by firms established by founders from universities and government labs. Sleeper is able to determine the department of origin of 27 of the 36 university founders. She finds that the firms founded by the 12 scientists coming from physics departments have a higher survival rate than other firms, including those that originated as laser spin-offs.

they employ is the fact that articles list names of authors and addresses for authors. Hicks uses the address information to classify by sector each of the 376,226 papers indexed in the Science Citation Index (SCI) with a United Kingdom address published during the 11-year interval of 1981–1991. A related methodology is used by CHI to analyze publishing patterns for *Science and Engineering Indicators.*

Prominent trends include the following:

- Increased multiple authorship. Although the growing importance of coauthorship has been observed for a long time,[25] Hicks and Katz (1996b) document that the increase is attributable to growth in papers with four or more authors; the proportion of papers with one or two authors was in decline during the period of study while the share of papers with three authors remained steady. Such detail is not readily available for U.S. articles but the incidence of multiauthored papers has risen from 45 to 56 percent during the period 1981–1995 (National Science Board, 1998, Table 5-53).[26]
- Increased intersectoral collaboration. Hicks and Katz find that during the 1980s U.K. papers published by authors located at a single institution did not grow while the number published by authors working in more than one institution rose steadily. Specifically, by the end of the period of study the proportion of collaborative papers rose from 28 percent of all U.K. papers to 41 percent. Increased intersectoral collaboration is occurring in the United States as well (National Science Board, 1998, pp. 5-38).
- Increased collaboration between industry and universities. During the 1981-1991 period, the percent of U.K. industry papers that included a university address rose from slightly less than 20 percent to slightly less than 40 percent (Hicks and Katz, 1997a, p. 138). Similar trends are occurring in the United States (National Science Board, 1998, pp. 5-38).
- Change in publishing patterns within industry. Hicks and Katz document that manufacturing is not the only sector of the U.K. industry that publishes. Indeed, taking the *Times* 1000 companies as their base they find that the only sector in which all companies (in this case 10) published during the period was "water." Viewed from the perspective of the percent of all industrial publications, several nonmanufacturing sectors—e.g. oil, gas and nuclear fuels, engineering, electricity—produce a sizeable pro-

[25] de Solla Price (1986) analyzed the number of authors on papers listed in Chemical Abstracts from 1900 to 1960 and found that the proportion of single-authored papers began decreasing in the 1920s.

[26] Strickly speaking the comparison is for the period 1981–1985 and 1991–1995.

portion of industrial publications (Hicks and Katz, 1997a, p. 32). In the United States industry publications almost doubled between 1981 and 1995 in clinical medicine and tripled in biomedical research. U.S. industry publications in physics, chemistry, technology, and mathematics all declined during the 1990s (National Science Board, 1998, pp. 5-38).

- Increased international collaboration. Hicks and Katz also find that for articles having at least one U.K. address the average number of countries per article increased during the period 1981 to 1991, from 1.17 to 1.25 (1996a, p. 390). The increase in U.S. scientists' participation in international collaborative research is seen by the fact that the proportion of articles with one or more U.S. addresses along with a non-U.S. address rose from 9 to 16 percent during the period 1981–1995 (National Science Board, 1998, Table 5-53).

Bibliometric research holds remarkable promise for using human resource data to study innovation if links can be made between biblio-metric information and data collected in such surveys as the SDR with files from funding agencies concerning the amount and source of research support. For example, if we were able to make such linkages we would gain insight into whether collaborations stem from attendance at the same graduate school, work with a dis-sertation advisor, work as a post doc, or work with a former employer. Understanding how these collaborative relationships are formed is crucial given the increasing importance of networks and the great difficulty faced by researchers in tracking informal relationships. Of course, not all informal relationships produce papers, but papers are one indication that a relationship exists.[27]

Patents

Patent applications in the United States include references to U.S. and foreign patents as well as "other" references, many of which are published articles. In recent years there has been guarded but growing interest in using these citations to study the process of innovation.[28] This interest is

[27] Zucker and Darby quote a manager as saying: "Copublishing is about as good an indicator as you can get of commonality of interests between [the company] and an academic collaborator. Although formal relationships are on a publicly available list, many relationships are not publicly acknowledged." The investigators continue, "In this and other fieldwork we have repeatedly validated the usefulness of linking academic scientists to firms by bibliometric research on patterns of co-publication... This concept of linkage is powerfully predictive of firm success when academic star scientists are involved." (Zucker and Darby, 1995, p. 22).

driven in part by the fact that the use of patent counts as an indicator of innovation has not lived up to expectations because of what Trajtenberg (1990, p. 172) refers to as "the enormous variance in their 'importance' or 'value.' " Weighting patent counts by the citations received in subsequent patent applications, however, provides a means of attributing value to the patent. Trajtenberg (1990) demonstrates that this weighted measure performs much better as an indicator of the value of innovations than does total patent count.[29] Hall, Jaffe, and Trajtenberg (1998) provide evidence that a citation-weighted patent measure is contemporaneously associated with market value. To date no one has attempted to link authors of cited and citing patents to analyze characteristics of the spillover process that could be learned through this human resource link.

Considerable information can also be gleaned from articles cited in patent applications.[30] Narin, Hamilton, and Olivastro (1997) use these cites to determine the origin of the basic science that underlies the patent. Their analysis shows that 73 percent of the references to published articles were to "public" science—that is science authored at academic, governmental, and other public institutions. They also find that the number of references to public science nearly tripled during the 6-year period studied. The research also indicates that NSF was the most widely acknowledged support agency in cited chemistry, physics, and engineering papers. NIH was the most widely cited in biomedical papers.

More recently, Narin collaborated with Deng and Lev (Deng et al., 1999) to demonstrate how the use of patent citation information adds to our understanding of the performance of firms in capital markets using such measures as stock returns and market-to-book ratios. Three mea-

---

[28] Hall, Jaffe, and Trajtenberg (1999, p. 5) discuss the substantial "noise" in patent citation data. In addition to cites that are included in the patent document because of the knowledge linkage, citations are also made for legal reasons. There are also what they call "after the fact" citations—those added to the document after the actual invention and what they call "teaching" cites, those that everyone considers basic. In addition, the patent examiner may also require the addition of relevant citations to "further bound the scope of intellectual property rights conferred by the patent, even though the inventor may not have been aware of the patent to which the citation is added."

[29] The dependent variable in the Trajtenberg study is the gains accruing to the representative consumer as well as the total gains to all consumers. Trajzenberg focuses his study on the CT scanner industry.

[30] Deng, Lev, and Narin (1999 p. 21) report that a typical U.S. patent cites about eight earlier U.S. patents and one to two foreign patents. In addition, the typical patent cites one or two nonpatent references, the majority of which are science references. In recent years, there has been a steady increase in the number of patents referenced and the citations to science, the latter having grown from an average of .3 per patent application in 1985 to 2.0 in 1997 with a tripling of citations to U.S. scientific papers in the 1988–1994 span.

sures derived from the citations are constructed: one measures the importance of the patent through the use of forward citations to the patent; another measures the link between the patent and science by measuring the number of references to scientific papers; and a third indicator measures the median age of patents referenced in the application and is developed as a measure of how quickly a technology is evolving.

The authors find the three patent measures as well as a measure of patent counts to be significantly related to market-to-book value with the expected sign. Of particular interest is the fact that none of the three patent variables contain information that is obtainable in the companies' financial reports. The stock return results behave somewhat similarly, although the association between patent attributes and subsequent stock returns is generally weaker than the association between patent attributes and market-to-book value.

The patent studies are indicative of the richness of insights that can be gained by using existing information in creative ways. Additional insights could undoubtedly be gained by *linking* these data with human resource data. The funding link established by Narin is suggestive, but substantially more could be learned if we were able to make human resource links. For example, one could establish the educational origins of individuals authoring highly cited patents and more readily determine sources of funding for the research of cited articles by using the name as the link between publication and funding agency. Similar knowledge could be gained if we were able to establish human resource links on a regular basis with publication and citation information, as collected by ISI.

SECTION VI. CONCLUSION

Four broad changes have occurred in the structure and organization of innovative activity in the United States: (1) decreased role for federal funding of R&D; (2) a change in the industrial distribution of innovative activities; (3) a shift of resources toward development activities and away from basic research; and (4) a change in the organization of research. The latter change reflects an increased reliance on external R&D, increased collaboration in the development of new products and processes, a decentralization of in-house R&D activities, and the movement of innovative activities to functions in the firm typically not thought of as being drivers of innovation. These changes mean that traditional indicators of R&D as well as the traditional unit of analysis, the firm, are less relevant to the study of innovation than they once were.

Here we have explored how human resource data can be used to illuminate patterns of innovation and resource utilization and, perhaps more importantly, what could be learned if the data that we already have were linked, providing insights into networks and the avenues by which collaborative ventures are formed and knowledge moves across boun-daries. Such links would provide a clearer picture of the process of innovation and the causes of economic growth. A nontrivial benefit from such research is that it offers the possibility of providing a clearer under-standing of how the investments of government and the nonprofit sector contribute to economic growth.

Our preliminary investigation shows that certain changes in the structure and organization of innovative activity can readily be seen by using HR data. For example, the deployment data show a change in industrial mix that R&D data support but fail to fully capture. These trends could be more clearly discerned if the firm address were carefully coded.[31] We also find that the HR data provide insight into the movement of innovative activities to non-R&D functions in the firm. The HR data also allow us to see how the geographic distribution of scientists and engineers changes over time.

Much more could be learned if we were able to link this HR data to other already established databases. For example, we know a great deal about firms but have made no effort to link information on firms to HR data collected by such agencies as NSF. Publication and citation data are readily available going back for a number of years. In addition to containing author name and journal name, the SCI data also include address information for authors. These data have rarely been linked to HR data. The lack of linkage means that we have forfeited information on the nature of collaboration between industry and academe as well as the way in which firms based on new technologies are founded. A similar case can be made for patent data. We have long used patents as an indicator of innovation but only recently have become interested in using patent citations to learn about the science linkages of inventions as well as to measure the importance of the patent. In the past linkages such as these appeared to be luxuries; but as boundaries and roles continue to blur they cease to be luxuries. In a world where the process of innovation is radically changing, valuable information is being lost by our failure to create and analyze HR linkages in a systematic way.

[31] NSF reports that in many instances the employer address data (city, state, zip) are missing and that the employer name is reported in unclear acronyms. This suggests that an effort would need to be made to get cleaner information from respondents if matching were to occur in the future.

REFERENCES

Acs, Zoltan J., David Audretsch, and Maryann Feldman. 1992. "Real Effects of Academic Research: Comment." *American Economic Review,* 82(l): 363-67.

Audretsch, David, and Paula Stephan. 1996. "Company-Scientist Locational Links: The Case of Biotechnology." *American Economic Review,* 86(3): 641-52.

______. 1999. "Knowledge Spillovers in Biotechnology; Sources and Incentives." *Journal of Evolutionary Economics,* 9(1): 97-108.

Badaracco, Joseph. 1991. *The Knowledge Link.* Boston: Harvard Business School Press.

de Solla Price, Derek. 1986. *Little Science, Big Science and Beyond.* New York: Columbia University Press.

Deng, Zhen, Baruch Lev, and Francis Narin. 1999. "Science and Technology as Predictors of Stock Performance." *Financial Analysts Journal,* 55(3): 20-32.

Finn, Michael, Leigh Ann Pennington, and Kathryn Hart Anderson. 1995. "Foreign Nationals Who Receive Science or Engineering PhDs from U.S. Universities: Stay Rates and Characteristics of Stayers." Oak Ridge, TN: Oak Ridge Institute for Science and Education.

Friar, J., and Horwitch, M. 1986. "The Emergence of Technology Strategy: A New Dimension of Strategic Management," in Mel Horwich, ed. *Technology in the Modern Corporation: A Strategic Perspective.* New York: Pergamon Press.

Hagedoorn, John. 1993. "Understanding the Rationale of Strategic Technology Partnering." *Strategic Management Journal,* 14: 371-385.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. 1998. "Market Value and Patent Citations: A First Look." Working paper, University of California at Berkeley.

______. 1999. "Market Value and Patent Citations: A First Look." Unpublished paper.

Hamel, Gary. 1991. "Competition for Competence and Inter-Partner Learning within International Strategic Alliances." *Strategic Management Journal,* 12: 83-103.

Hicks, Diana, and J. Sylvan Katz. 1996a. "Science Policy for a Highly Collaborative Science." *Science and Public Policy,* 1: 39-44.

______1996b. "Where is Science Going?" *Science, Technology & Human Values*, 21(4): 379-406.

______1997a. "The Changing Shape of British Industrial Research." STEP Special Report no. 6.

______1997b. "A National Research Network Viewed from an Industrial Perspective," *Revue D'Economie Industrielle*, 79(l): 129-142.

Jaffe, Adam B. 1989. "Real Effects of Academic Research." *American Economic Review*, 79(5): 957-70.

Lev, Baruch. 1999. "R&D and Capital Markets." *Journal of Applied Corporate Finance*, 11(4): 21-35.

Levin, Sharon G., and Paula E. Stephan. 1991. "Research Productivity Over the Life Cycle: Evidence for Academic Scientists." *American Economic Review,* 81: 114-32.

Leyden, D.P., and A.N. Link. 1999. "Federal Laboratories as Research Partners." *International Journal of Industrial Organization*, 17(4): 575-592.

Merrill, Stephen A., and Ronald S. Cooper. 1999. "Trends in Industrial Research and Development: Evidence from National Data Sources" in National Research Council. *Securing America's Industrial Strength.* Washington, D.C.: National Academy Press.

Mowery, David C. 1999. "America's Industrial Resurgence: How Strong, How Durable?" *Issues in Science and Technology*, 15(3): 41-48.

Narin, Francis, Kimberly S. Hamilton, and Dominic Olivastro. 1997. "The Increasing Linkage between U.S. Technology and Public Science." *Research Policy,* 26(3): 317-330.

National Science Board. 1998. *Science & Engineering Indicators—1998*. Arlington, VA: National Science Foundation.

Nerad, Maresi, and Joseph Cerny. 1999. "Postdoctoral Patterns, Career Advancement, and Problems." *Science,* 285: 1533-1535.

Powell, Walter W., and Jason Owen-Smith. 1998. "Universities and the Market for Intellectual Property in the Life Sciences." *Journal of Policy Analysis and Management*, 17(20): 253-277.

Powell, Walter W., Kenneth W. Koput, Laurel Smith-Doerr, and Jason Owen-Smith. 1998. "Network Position and Firm Performance: Organizational Returns to Collaboration in the Technology Industry." Unpublished paper.

Saxenian, Annalee. 1994. *Regional Advantage.* Cambridge, MA: Harvard University Press.

Sleeper, Sally. 1998. *The Role of Firm Capabilities in the Evolution of the Laser Industry. The Making of a High-Tech Market.* PhD Dissertation, Department of Social and Decision Sciences, Carnegie Mellon University.

Stephan, Paula. 1999. "Capitalizing on Human Capital: The Case of Biotechnology." Unpublished paper.

Stephan, Paula, and Stephen Everhart. 1998. "The Changing Rewards to Science: The Case of Biotechnology." *Small Business Economics,* 10: 141-151.

Stephan, Paula, and Sharon Levin. 1991. *Striking the Mother Lode in Science: The Importance of Age, Place, and Time.* New York: Oxford University Press.

Trajtenberg, Manuel. 1990. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *Rand Journal of Economics,* 21(l): 172-187.

Zucker, Lynne, and Michael Darby. 1995a. "Present at the Revolution: Transformation of Technical Identity for a Large Incumbent Pharmaceutical Firm after the Biotechnological Breakthrough." NBER, Working Paper No. 5243.

______. 1995b. "Virtuous Circles of Productivity: Star Biochemists and the Institutional Transformation of Industry." NBER, Working Paper No. 5342.

_______.1997. "Academic Scientists-Entrepreneurs and Commercial Success in Biotechnology." Unpublished paper.

Zucker, Lynne, Michael Darby, and Marilyn Brewer. 1994. "Intellectual Capital and the Birth of U.S. Biotechnology Enterprises." National Bureau of Economic Research, Working Paper No. 4653.

Zucker, Lynne, Michael Darby, and Maximo Torero. 1997. "Labor Mobility From Academe to Commerce." NBER, Working Paper No. 6050.