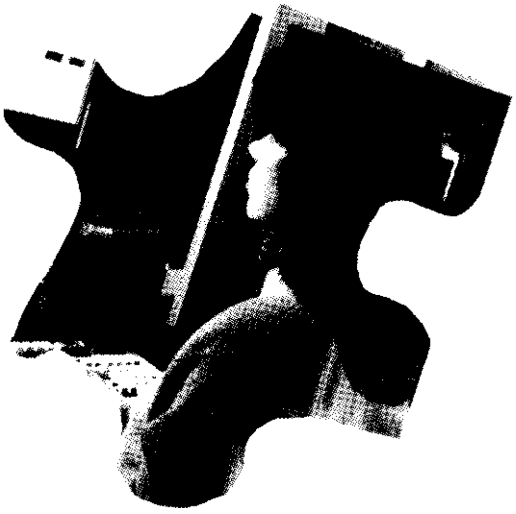
CHAPTER 8

Data on the Federal Research System



Contents

Boxes

Figures

Tables

CHAPTER 8

Data on the Federal Research System

> *The measurement process is also inherently limited by the inevitable human selection of both the phenomena to be measured and the type of data considered relevant to the purpose of the measurement effort. . . . For measuring an area as little understood as the science and technology enterprise, multiple models are needed to insure that as wide a spectrum of phenomena as possible is included.*
>
> U.S. General Accounting Office[1]

Introduction

While this report has characterized the Federal research system as it enters the 1990s, its mandate was broader. OTA was asked what data and analytical tools would be useful in describing the research system. Preceding chapters have drawn on much data. However, there are many areas in which additional information would be welcome. Data, in short, are an *issue in* Federal research policy, especially their form, gaps, and uncertainties. For example, OTA has discussed:

- variable definitions of 'scientists' and ' 'engineers" that can result in radically different estimates of their numbers (chapter 1);
- problems with using different deflators to calculate constant dollar trends in research funding (chapter 2);
- potential comparisons between congressionally earmarked and peer-reviewed projects (chapter 3);
- lack of information on how agencies process research proposals prior to awards (chapter 4);
- problematic estimates of research expenditures in megaprojects (chapter 5);
- need for comparative cost-accountability data, by institution and source, on research expenditures (chapter 6); and
- lack of baseline information on the Nation's research work force, as opposed to all scientists and engineers (chapter 7).

Data collected on certain aspects of the Federal research system—sources and dollars spent for research, academic degrees awarded, facilities and instruments, and various outcome measures such as publications and citations—are extensive.

In other areas, however, data are scarce, for example, details on the *research* work force (as opposed to the total science and engineering work force), or what proportion of investigators-across fields and agencies—are supported by Federal funds. Also, compared to the National Science Foundation (NSF) and the National Institutes of Health (NIH), the other research agencies devote few resources to internal data collection. Consequently, most analysis and research decisionmaking must draw conclusions from the NSF and NIH data systems. Since these agencies represent only part of the spectrum of research supported by the Federal Government, these analyses may omit key results and trends at other agencies, or skew findings toward biomedicine or academic research.

Furthermore, it is not clear how available agency data are used to inform decisionmaking, as some challenge current policy assumptions and others are reported at inappropriate levels of aggregation. For example, while there is much attention paid to the rising cost of instrumentation and facilities, indirect and personnel costs are rising at faster rates and account for larger shares of Federal expenditures. In this case, the issue is not information, but what can be done with it by decisionmakers.[2]

In this chapter, OTA first summarizes the data that are currently available. Table 8-1 lists the new data that OTA gathered and examined for this report.

---

[1] U.S. General Accounting Office, Science *Indicators: Improvements Needed in Design, Construction, and Interpretation*, PAD-79-35 (Washington, DC: 1979), pp. 5-6.

[2] A distinction is made throughout this chapter between "decisionmakers" and "policymakers." The former comprise a considerably larger population especially within the Federal research agencies; the latter are found at the very top of those agencies, as well as in the Office of Management and Budget, the Office of Science and Technology Policy, and Congress. Data speak to research decisionmakers at all levels, some of whom are not responsible for policies.

Table 8-l—Summary of OTA Data Collection and Analysis on Federally Funded Research

Description	Methods of collection	Subject
Original data collection and analysis:		
Federal agency analysis[a]	Interviews, site visits, and document review	Priority setting and funding allocation
University case studies[b]	Interviews and site visits	Rising research costs and responsiveness to changing priorities
Bibliometrics[c]	Citation analyses	"Hot" fields, related fields, university comparisons, and other indicators
Analysis of SEI[d]	interviews and document review	Evolution of SEI volumes, data presentation, and future analysis
Researchers' views[e]	Surveys	Sigma Xi members' perceptions of Federal research funding issues
Secondary data analysis:		
Research cost Comparisons[f]	NSF, NIH, and other datasets	Rising costs of research
Country surveys[g]	interviews and document review	Priority setting, funding allocation, and research evaluation in other countries
Congressional earmarking	Budget analysis and document review	Budget information on congressional funding and definitions of "earmarks"
Rhetorical analyses[-]	Document review	Historical analysis of research decisionmaking by different branches of government and goals of different ideological groups
Research evaluation[j]	interviews and document review	Post-1 985 developments in research evacuation in the United States and abroad
Analysis of Science Policy Task Force hearings[k]	Document review	Analysis of House hearings on science policy, 1985-87

[a]See box 4-A, chapter 4.
[b]See box 6-B, chapter 6.
[c]Henry Small and David Pendlebury, "Federal Support of Leading Edge Research," OTA contractor report, February, 1989; and Henry Small, "Bibliometrics of Basic Research, " OTA contractor report, July 1990. See appendix F for information about how to obtain the latter report and all other OTA contractor reports listed below.
[d]Susan Cozzens, "Science Indicators: Description or Prescription?" OTA contractor report, July 1990.
[e]John Sommer, "Researcher Perspectives on the Federal Research System," OTA contractor report, July 1990.
[f]Kathi Hanna, "Federal Funding of Basic Research," OTA contractor report, November, IWO; and Harvey Averch, "Analyzing the Costs Of Federal Research," OTA contractor report, August 1990.
[g]Ron Johnston, "Project Selection Methods: International Comparisons," OTA contractor report, June 1990.
[h]James Savage, "Academic Earmarks and the Distribution of Federal Research Funds: A Policy Interpretation," OTA contractor report, July 1990.
[i]See Mark Pollock, "Basic Research Goals: Perceptions of Key Political Figures," OTA contractor report, June 1990; and David Birdsell and Herbert Simons, "Basic Research Goals: A Comparison of Political Ideologies,"OTA contractor report, June 1990.
[j]Harvey Averch, "Policy Uses of 'Evaluation of Research' Literature," OTA contractor report, July 1990.
[k]Patrick Hamlett, "Task Force on Science Policy: A Window on the Federal Funding and Management of Research," OTA contractor report, October 1990.

KEY: SEI = *Science & Engineering Indicators*; NSF= National Science Foundation; NIH = National Institutes of Health.

SOURCE: Office of Technology Assessment, 1991.

Second, OTA suggests additional information that could be collected, concentrating in areas of policy relevance and on data that are amenable to manipulation in the aggregate by Congress and the executive branch (especially the Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB)), and at less aggregated program and project levels within the research agencies. The emphasis is on the analysis and presentation of data for monitoring changes in the Federal research system. Finally, OTA considers the utility of data for decisionmaking, revisiting the problems of evaluating research projects (and updating conclusions of a previous OTA study).

Information is not cost-free. While it can illuminate the operation of the Federal research system, for all participants and at many levels, the purpose is not to generate needless paperwork and impose new reporting requirements on the agencies. What may be appropriate for decisionmakers, in fact, is less information, not more, along with better measures and methods of applying and coordinating it.

What Data Are Available on the Federal Research System?

Many organizations collect and analyze data on the research system. First and foremost, is NSF, with its numerous surveys, reports, and electronic data systems that are publicly available. Other sources include the other Federal research agencies; the National Research Council (NRC); the Congressional Research Service (CRS); professional societies, especially the American Association for the Advancement of Science (AAAS); and other public and special interest groups.[3]

Together these databases and analyses provide a wealth of information: time series on the funding of research and development (R&D); expenditures by R&D performer (e.g., universities and colleges, industry, Federal laboratories), by source of funds, and by type (basic, applied, development); numbers of students who enroll in and graduate with degrees in science and engineering (s/e); characteristics of precollege science and mathematics programs and students in the education pipeline; and size, sectors of employment, and activities of the s/e work force (especially Ph.D.s in academia). Detailed analyses of the Federal budget by research agency are available each year, and impacts on specific disciplines and industries can often be found.

NSF publishes many annual or biennial reports. These reports summarize budget data from the Federal agencies, academic R&D (which is covered extensively, as academia is NSF's primary client), research at the Federal laboratories, funding and performance of research by industry, academic equipment and instrumentation expenditures, international comparisons, geographic distributions of R&D funds, and other topics. NSF also publishes detailed information on students, degrees awarded, employment by sector, and the people who perform research. Finally, NSF issues many individual reports on specific topics either requested by Congress or of particular interest to the scientific community.

Photo credit: National Aeronauts and Space Administration

An astronaut spins liquid in zero gravity aboard the Space Shuttle Columbia to test the separation of bubbles from the liquid. Research can take place in many different settings.

Certainly the most visible compendium of data on the research system is the biennial report, *Science& Engineering Indicators (SEI)*, issued since 1973 by the National Science Board (NSB), the governing

[3]For a summary of major databases on science and engineering (individuals and institutions), see National Research Council, *Engineering Personnel Data Needs for the 1990s* (Washington, DC: National Academy Press, 1988), app. A-2.

body of NSF.[4] At 1976 hearings,[5] NSB chairman Norman Hackerman traced the origins of SEI to a congressional mandate to NSB for its annual report, which was to focus on ''. . . the status and health of science and its various disciplines (including) an assessment of such matters as national scientific resources, . . . progress in selected areas of basic scientific research, and an indication of those aspects of such progress which might be applied to the needs of American society. ' '[6] From the outset, then, the SEI project aspired to measure and evaluate the results of federally supported R&D.

Table 8-2 lists eight broad categories of data that have appeared in SEI, including impacts and assessments, resources, scientific performance, economic performance, international contacts, cross-sectoral linkages, literacy, educational pipeline, and scientific work force. Table 8-3 shows the distribution of tables among data types in the nine SEI reports. Even this broad-brush picture reveals a highly dynamic volume. Over the years, 79 distinct subcategories of data have appeared, about one-half in the original 1972 volume and about one-half added later. The categories of international and cross-sectoral contact, literacy, and pipeline show steady patterns of expansion in types of data. Resources, impacts and assessments, and scientific performance indicators have been stable, with some new types added. The economic performance indicators show high turnover—many categories added and some dropped. Publication and citation measures are still the main forms of scientific performance data.[7]

SEI stands as the most comprehensive look at the research system that is currently available. Some find fault with the volume, however, because it is based on an input/output model of science (i.e., "people and money enter the system, research comes out"), which is thought to be simplistic, omitting quantitative (and qualitative) measures of the ***process*** of research.[8] Others criticize SEI for its concentration on academic or academically based research and lack of emphasis on the research-technology interface.[9]

Each year, CRS and AAAS publish perhaps the most comprehensive and widely read compilations of Federal R&D spending (the former focuses on appropriations, the latter on the proposed "R&D budget"). These documents help to interpret, by placing into an historical frame, the appropriations bills signed into law by the President. Various professional societies, e.g., the American Chemical Society, also compile surveys of R&D spending, salaries, and employment opportunities that are of particular interest to their constituencies. In addition, the American Council on Education, the Council of Graduate Schools, and the Association of American Universities publish annual and occasional reports that characterize trends in research university expenditures, administrator and faculty

---

[4] The following discussion of *Science and Engineering Indicators* (SEI) is based on Susan E. Cozzens, "Science Indicators: Description or Prescription?" OTA contractor report, July 1990. Available through the National Technical Information Service, see app. F. Note that SEI was named *Science Indicators until 1987. Science& Engineering Indicators* builds on data collected, published, and issued in many other reports by the Science Resources Studies Division of the National Science Foundation.

[5] The timing of these hearings was important in the development of the Indicators series. The 1972 volume had been resembled in the course of a few months by one staff person working with unenthusiastic and energetic Board committee. After a stormy and uncertain process of approval both within the Science Board itself (who could not agree on how the numbers should be interpreted) and at the Office of Management and Budget and the White House (who thought it presented administration policy in too unfavorable a light), the volume appeared amidst considerable fanfare in the science and general press (as reported by Robert Brainard, the National Science Foundation staff member who prepared the first report). See U.S. Congress, House Committee on Science and Technology, Subcommittee on Domestic and International Scientific Planning and Analysis, *Measuring and Evaluating Results of Federally Supported Research and Development: Science Output Indicators—Part I. Special Oversight Hearings, 94th* Cong., May 19 and 26, 1976 (Washington, DC: U.S. Government Printing Office, 1976).

[6] Ibid., p. 7.

[7] The bibliometric database has added more to the categories of international and cross-sectoral contacts than it has to measures of scientific performance.

[8] For other volumes that address these issues, see National Academy of Sciences, *The Quality of Research in Science: Methods for Postperformance Evaluation in the National Science Foundation (Washington,* DC: National Academy Press, 1982); Y. Elkana et al., *Toward a Metric of Science (New* York, NY: Wiley, 1977); and H. Zuckerman and R.B. Miller (eds.), "Science Indicators: Implications for Research and Policy," *Scientometrics*, vol. 2, October 1989, special issue, pp. 327-448.

[9] Cozzens, op. cit., footnote 4. Because *Science and Engineering Indicators* (SEI) should reflect analytical advances in characterizing science and technology, provision could be made for the support of relevant research communities outside of the National Science Foundation (NSF). According to NSF's Carlos Kruytbosch (personal communication, December 1990), at the very least ". . . biennial post-publication workshops to evaluate the SEI report are workable and could be productive. "

Table 8-2—Categories of Data in Science& *Engineering Indicators*

International contacts
Cross-national citations and coauthorships, publishing in foreign journals, participation in international scientific congresses, employment plans of foreign students, U.S. students and academics going abroad.

Cross-sectoral linkages
Citations from patents to the scientific literature, cross-sectoral coauthorship, cross-sectoral citation, mobility between sectors, university patenting.

Economic performance
Patents, trade and trade balances, productivity measures, global investments, innovation indicators, high-technology business sector, venture capital.

Impacts and assessments
Public views on allocation of resources for science, judgments of benefit and harm from science and technology, prestige of scientists, expectations of scientific advances and problems caused by science, differences between the attentive and general public.

Literacy
Enrollments in science and mathematics, course content and testing requirements, achievement and test scores, teacher characteristics and activities, public understanding of scientific concepts, public use of technologies, student attitudes toward science and technology.

Pipeline
College and graduate school enrollments in science, engineering, and mathematics; degrees; test scores and other quality measures; preferences and plans of high school and college students; sources of student support.

Resources
Expenditures and obligations, special research resources, instrumentation and facilities.

Scientific performance
Publication and citation counts, Nobel Prizes.

Work force
The science and engineering work force: comparative measures, demographic characteristics, career variables, sources of support, technicians, stock and flow analysis.

SOURCE: Susan Cozzens, Science Indicators: Description or Prescription?" OTA contractor report, July 1990.

issues, and Federal support for education and research.[10]

Recently, the Government-University-Industry Research Roundtable of the National Academy of Sciences, with data compiled by NSF's Policy Research and Analysis Division, provided much useful analysis on the state of academic R&D and changes since the early 1960s.[11] In addition, NRC periodically publishes reports on sectors of the research system and on the availability of data to characterize the system.[12] These publications provide a basis for understanding the Federal research system. But even with each of these organizations devoting significant resources to the collection of information, better data are needed to guide possible improvements of the system.[13]

What Data Are Needed?

Recognizing that data collection is often very difficult, and certainly time consuming, OTA concentrated on notable gaps in the empirical baseline. One overarching problem is that comparable data

[10] A monthly compendium that announces and annotates new reports containing data and analysis on trends in science and engineering is *Manpower Comments,* published by the Commission on Professionals in Science and Technology, a participating organization of the American Association for the Advancement of Science.

[11] Government-University-Industry Research Roundtable, *Science and Technology in the Academic Enterprise: Status, Trends, and Issues* (Washington, DC: National Academy Press, October 1989).

[12] For example, see National Research *Council, Surveying the Nation's Scientists and Engineers: A Data System for the 1990s* (Washington, DC: National Academy Press, 1990). Under multiagency support, the National Research Council collects, analyzes, and disseminates information on Ph.D. recipients. For a statement of its cross-cutting role, see National Academy of Sciences, *The National Research Council: A Unique Institution* (Washington, DC: National Academy Press, 1990).

[13] These efforts must also be seen in the context of the massive Federal data system. The components most relevant to research are the data series compiled and reported by the Census Bureau, the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the National Center for Education Statistics. The point is illustrated by calls for ways to measure how many people who aspire to attend college actually enroll. In the words of one sociologist: "We care to know on a month-to-month basis what the unemployment rate is. I think we ought to care to know on at least a year-to-year basis what the rate of access to higher education is. " Quoted in Thomas J. DeLoughry, "U.S. Asked to Set Student-Aid Goals for Poor and Minority Students," *The Chronicle of Higher Education, vol. 37, No. 20,* Jan. 30, 1991, p. A20.

Table 8-3-Trends in Distribution of Data Among Categories in *Science &* Engineering Indicators: 1972-88

Number of tables

	1972	1974	1976	1978	1980	1982	1984	1986	1988	Total
Resources	38	63	57	58	41	52	49	47	60	465
Work force	18	37	35	35	54	37	29	38	30	313
Economic performance	9	31	47	22	40	37	20	35	45	286
Impacts and assessments	21	14	21	0	39	35		29	18	196
Pipeline	23	13	9	3	1	14	19	29	26	140
literacy	1	0	0	0	1	4	14	46	54	120
Scientific performance	2	10	11	10	11	15	7	4	7	77
International contacts	0	2	8	10	11	11	11	9	8	70
Cross-sectoral contacts	0	3	0	3	8	7	6	12	7	46
Nonindicators	0	0	0	0	1	1	1	0	3	6
Total	112	173	188	141	207	213	178	249	258	1719

Percent of total

	1972	1974	1976	1978	1980	1982	1984	1986	1988	Total
Resources	34	36	30	41	20	24	28	19	23	27
Work force	16	21	19	25	26	17	16	15	12	18
Economic performance	8	18	25	16	19	17	11	14	17	17
Impacts and assessments	19	8	11	0	19	16	11	12	7	11
Pipeline	21	8	5	2	0	7	12	12	10	8
Literacy	1	0	0	0	0	2	8	18	21	7
Scientific Performance	2	6	6	7	5	7	4	2	3	4
International contacts	O	1	4	7	5	5	6	4	3	4
Cross-sectoral contacts	O	2	0	2	4	3	3		3	3
Nonindicators	O	0	0	0	0	0	1	:	1	0
Total	100	100	100	100	100	100	100	100	100	100

NOTE: Each table, text, or appendix is counted once.

SOURCE: Susan Cozzens, "Science Indicators: Description or Prescription?" OTA contractor report, July 1990.

associated with the research operations of *all the* Federal agencies are lacking. NSF and NIH conscientiously log data on what proportion of proposals submitted to them are awarded finding, the number of researchers they support, expenditures by categories of project budgets (e.g., indirect costs and personnel), and other dimensions related to management of their research programs. However, other agencies collect only R&D budgetary information, primarily in response to OMB requests and NSF surveys of research conduct. Much more data could be collected on research funding and performance in these agencies. In particular, further information could be collected on proposal submissions as well as awards, research expenditures by line items of the budget (requested and expended), and the size and distribution of the research work force that is supported. Comparable data from the agencies are important for decisions that span agencies or broad segments of the scientific community. With the data that are currently available, Congress and other Federal policymakers risk overgeneralizing from what is known about research performance that is supported by NSF or NIH

Some advocate that NSF should be the sole agency to centralize and standardize the analysis, especially since NSF has the mandate to collect R&D funding data from the other research agencies.[14] However, OTA found that the research agencies are sufficiently diverse in their organization and finding structures to create difficulties for any outside agency to translate data in comparable ways.[15] For example, breakdowns of R&D into basic research, applied research, and development are

[14]See Commission on Professionals in Science and Technology, *Measuring National Needs for Scientists to the Year 2000, report* of a workshop, Nov. 30-Dec. 1, 1988 (Washington, DC: July 1989).

[15]Although problems may exist with definitions, compliance by the reporting organizations with whatever definitions are used is also an issue. The advantage of an interagency mechanism, such as the Office of Science and Technology Policy's Federal Coordinating Council for Science, Engineering, and Technology (FCCSET) committees, is its place in the Federal hierarchy: the agencies are likely to be responsive to requests for "crosscutting" information where budgets are at stake. The FCCSET Committee on Physical, Mathematical, and Engineering Sciences, for example, currently has a "structure of science' activity that includes the solicitation of data from the research agencies similar to those sought in this OTA study.

very difficult to measure and often judgments are made after-the-fact. NSF fields a survey to all Federal R&D agencies asking for detailed estimates of spending in various categories. Because of problems with applying definitions and with converting agency accounting of research dollars into the separate categories, however, many of the agencies claim that it is impossible to provide accurate answers to the NSF survey.

In 1989, NSF continued its effort to develop a better taxonomy of the research it funds.[16] "Fundamental,' "directed," and "development" seemed to be the preferred categories, though some programs found it difficult to translate currently supported projects into these three categories. Unfortunately, any taxonomy would suffer from arbitrary divisions of research topics among categories. Also, "basic" and "applied," or similar definitions, are rarely used by managers to allocate monies; rather these distinctions are most important to the researchers who perform the research, since basic research is synonymous with enhanced investigator discretion over research directions, while applied research is often associated with the attainment of specific objectives.[17] Consequently, basic and applied divisions are less important for decisions that concern specific programmatic goals; however, they are quite important to decisions about the science base supported by the Federal Government.

Enhanced data collection at each agency would help NSF fulfill its data mandate, and advance development of comprehensive research strategies, especially programs that span agencies.[18] Other data that could be very useful fall into four categories: 1) research monies—how they are allocated and spent; 2) personnel-characteristics of the research work force; 3) the research process-how researchers spend their time and their needs (e.g., equipment and communication) for the conduct of research; and 4) outcomes—the results of research.

Research Monies

While the data collected on research sponsored by the Federal Government are abundant, information on research expenditures is not. In particular, direct and indirect costs in all sectors of the research system supported by the Federal Government could be monitored.[19] NSF and NIH have collected longitudinal data on research expenditures in individual investigator grants, but complementary data are needed on expenditures in Federal and industrial laboratories, research supported by other agencies, and on other types of research groups and cooperative ventures such as centers and university-industry collaborations. These data would help to monitor fluctuations in research expenditures. At present, predictions of future spending merely extrapolate from the gross totals disaggregated by sector, while individual components of the budgets may be increasing or decreasing relative to overall trends. These data would be especially helpful for revising estimates of start-up and operating costs in science megaprojects.

Another measure that would refine the knowledge of research expenditures would be breakdowns by field. (This is available for some academic research disciplines and Federally Funded Research and Development Centers only.) Many claims are made about the cost requirements of specific fields. For instance, research in some physics specialties is inherently more expensive than in others, because of the equipment required by research groups. At

16NSF Task Force on Research and Development Taxonomy, "Final Report,' unpublished document, 1989. For an earlier effort, see National Science Foundation *Categories of Scientific Research (Washington DC: 1979).*

17See Harvey Averch, "The Political Economy of R&D Taxonomies," *Research Policy*, forthcoming 1991; and Richard R. Ries and Henry Hertzfeld, "Taxonomy of Research: Test of Proposed Definitions on the NSF Budget," unpublished document, n.d.

18For e-pie, the National Institutes of Health sets aside 1 percent of its research budget for research evaluation and internal analysis Of the investigators and programs it supports. The Department of Energy, the National Aeronautics and Space Administration, the Office of Naval Research, and the National Science Foundation have all conducted ad hoc inhouse evaluations of the research they support and the efficiency of the operations needed to select and manage various research portfolios (see below). For an example of agency-based research evaluation data that could be assembled in an ongoing way, see Daryl E. Chubin, "Designing Research Program Evaluations: A Science Studies Approach," *Science and Public Policy,* vol. 14, No. 2, April 1987, pp. 82-90.

19This monitoring is not the same as the auditing of cost data by category of expenditure, as mandated by Office of Management and Budget Circular A-21 and as conducted by the Department of Defense and the Department of Health and Human Services contract audit agencies. That is done for accountability purposes. Congress seeks better information on how investigators and their teams actually spend money in the course of executing federally funded research projects, which requires some demystification of university accounting schemes. For examples of studies of data audit methodologies, see the new quarterly journal, *Accountability in Research: Policies and Quality Assurance,* edited at the University of Maryland School of Medicine.

present, there are no means to evaluate these claims.[20] Yet for decisions that must balance the present and future needs of different sectors of the research system, such cost estimates and the trends associated with them could be very important.

Finally, data on how Federal agencies allocate monies within project budgets could be compiled. Agencies have much experience in negotiating budgets. Data would illuminate how judgments are made about specific categories of expenditure, e.g., in reducing "inflated' budget requests of investigators, imposing an artificial ceiling on equipment purchases, or adjusting allocations through NIH's practice of "downward negotiation." Since personnel costs have grown quickly compared with other research expenditures, financial analyses would be greatly enhanced by better personnel data.

Personnel

One of the most fundamental pieces of information on the research system is the size of the *research* work force, both in absolute numbers and as a fraction of all U.S. employed scientists and engineers. These numbers depend on how ''researcher" is defined.[21] While estimates exist of the rise in Ph.D. personnel employed in research universities, very little detailed data exist for industry or other sectors of the research system. Estimates of the *positions* held by Ph.D. personnel in academia are inadequate. Distinguishing nonfaculty research associates from postdoctoral fellows and full-time equivalent faculty is analytically important-and a nightmare to sort and track over time. Accurate estimates of the changing size of the research work force and how many are federally funded-and are seeking such support—would aid in measuring current and unfunded academic research capacity. In addition, accurate estimates of the numbers of researchers exiting the system would help to gauge the attractiveness of specific fields, as well as the category "science and engineering" relative to other occupations .22

Another trend that has been noted in this report, mostly with anecdotal evidence and inferences from analyses of expenditures, is the increasing size of research groups, both within the university structure and through Federal support of centers. This trend has policy implications for the cost of research, its interdisciplinary capabilities, and the changing demographics of the work force. It also reflects how researchers may spend their time. More data on "production units" in research and their dependence on Federal funding relative to other sources would augment enrollment, Ph.D. award, and work activity data. Changes in the structure of production units have also influenced the research process and the volume-and perhaps the character-of outcomes.[23]

Research Process

"How research is done" has evolved since the 1960s. In particular, the organization of research groups and the settings in which research is conducted have changed.[24] Data on the conduct of research would aid in understanding the opportunities and stresses on the Federal research system and in planning how the research system can adapt to changing conditions.

For instance, it is often claimed that researchers are spending much more time writing proposals, and that their research suffers as a consequence. No systematic data exist either to support or refute this claim. While it is in the interest of Federal sponsors for their grantees to spend as much time as possible in the conduct of research, investigators report that the increased competition for Federal funds compels

[20]For a recent effort to look comprehensively at Federal support, by agency and overtime, of one sector of one field, se American Chemical Society, Department of Government Relations and Science Policy, *Federal Funding of Academic Chemistry Research, FY 1980-FY 1988 (Washington* DC: November 1990).

[21]A ''researcher'' could be defined as anyone publishing a scientific paper (i.e., by authorship), possessing a Ph.D. (i.e., by credential), or working in a particular setting (i.e., by sector). Indeed, the problem of defining who is a 'scientist' also applies here. See Derek de Solla Price, *Little Science, Big Science (New* York, NY: Columbia University Press, 1963). Also see National Research Council, op. cit., footnote 3.

[22]For a discussion of methodological pitfalls associated with assessing, for example, characteristics of the Federal work force, see U.S. General Accounting Office, *Federal Work Force: A Framework for Studying Its Quality Over Time,* GAO/PEMD-88-27 (Washington DC: August 1988).

[23]The role of laboratory chief or team leader combines entrepreneurial and administrative/supervisory tasks. Both are essential to the funding and longevity of the productive research unit. On the emergence of the entrepreneurial role on campus, see Henry Etzkowitz, "Entrepreneurial Scientists and Entrepreneurial Universities in American Academic Science, ' *Minerva,* vol. 21, summer-autumn 1983, pp. 198-233.

[24]For a prophetic discussion see B.C. Griffith and N.C. Mullins, ' 'Coherent Social Groups in Scientific Change, " Science, vol. 177, Sept. 15, 1972, pp. 959-964.

Photo credit: U.S. Department of Agriculture

Researcher picks blueberries-95 percent of the varieties of blueberries in production today were developed by Department of Agriculture scientists. Research is performed in many settings.

proposal writing.[25] However, one might expect that as the size of academic research groups grows, principal investigators will spend more time seeking money to sustain their larger research teams and programs.[26] This phenomenon is similar to strategies in a law or consulting practice, where the addition of less senior associates leverages the effort of the more senior employees to spend more time marketing and winning projects for the firm. In the academic research community, entrepreneurial pursuits are very different from research and teaching, and the additional burden can be a source of stress for senior researchers.[27]

Many also claim that increasing time commitments required by research pursuits hamper the ability of faculty to meet their teaching responsibilities. Data on how faculty apportion their time have been unreliable. Ironically, self-reports in compliance with Federal accountability requirements tend to distort estimates of time spent on various work activities.[28] Since the Federal Government invests in the academic research system to maintain a strong instructional as well as knowledge-producing capability, shifts in the activities of researchers is of central concern.

Data are needed on how apprentice, junior (e.g., postdoctorates), and senior researchers spend their time on research (collecting data and analysis), proposal writing, teaching (classroom and one-on-one), travel, presenting results to scientific colleagues, and other pursuits.[29] Differences between time commitments in Federal, industrial, and academic settings could also be judged.[30]

More generally, data could be collected on changing equipment needs. The average lifetime of a scientific instrument has shrunk during the 1980s from an average of 7 years to less than 5 years.[31] Additional data could address such questions as: how does the reliance on equipment vary across fields? What happens to obsolete equipment? As

[25] For example, see Science: *The* End *of the* Frontier? a report from Leon M. Lederman, President-Elect to the Board of Directors of the American Association for the Advancement of Science (Washington DC: American Association for the Advancement of Science, January 1991).

[26] See D.E. Chubin and T. Connolly, "Research Trails and Science Policies: Local and Extra-Local Negotiations of Scientific Work," *Scientific Establishments and Hierarchies,* Sociology of the Sciences, Yearbook vol. 6, N. Elias (cd.) (Dordrecht, Holland: D. Reidel, 1982), pp. 293-311.

[27] For evidence on entrepreneurial behavior, see Karen Seashore Louis et al., "Entrepreneurs in Academe: An Exploration of Behaviors Among Life Scientists," *Administrative Science Quarterly, vol. 34, 1989,* pp. 110-131.

[28] See Donald Kennedy, "Government Policies and the Costs of Doing Research," *Science,* vol. 227, Feb. 1, 1985, pp. 480-484.

[29] For example, data could illuminate changing patterns of communication among scientific colleagues. With new communications technologies, such as electronic mail systems and computer networks, scientists have the ability to exchange data and ideas much more often. Is science becoming more collaborative (or competitive) due to these innovations? Do most scientists have access to these technologies? Are some at a disadvantage without them?

[30] Some clues derive from in situ laboratory studies of scientists, for example, Bruno Latour and Steve Woolgar, *Laboratory Life: The Social Construction of Scientific Facts* (Beverly Hills, CA: Sage, 1979); and Bruno Latour, *Science in Action: How To Follow Scientists and Engineers Through Society* (Cambridge, MA: Harvard University Press, 1987).

[31] See National Science Foundation, *Academic Research Equipment in Select Science/Engineering Fields: 1982-83 to 1985-86,* SRS 88-D1 (Washington DC: June 1988). As the National Science Foundation's Leonard Lederman (personal communication December 1990) points out, there is no information of average "equipment use rate," or what proportion of available time an instrument is in use.

communications and other technologies progress and the scientific community comes more to rely on them, these questions will increasingly impact Federal funding.

A final area of "process" on which data would be instructive are the standards for achieving various positions in the scientific community. Many claim that graduate students must publish more papers to be offered first jobs after receipt of the doctorate or completion of a postdoctoral fellowship. What are the average age, experience postcollege (in years), and publication records of new hires at research universities, and industrial and Federal laboratories? For other promotions? Such data would help the Federal Government to track the changing research labor market.[32]

Outcomes of Research

Because of the fundamental and elusive nature of research, measuring its outcomes—in knowledge and education—is very difficult.[33] The most elusive outcome is cultural enrichment-the discovery and growth of scientific knowledge. As OMB Director Richard Dar-man has said (speaking of the proposed Moon/Mars mission): "No one can put a price on uplifting the Nation. Research has resulted in many benefits to the Nation and is funded precisely because of those benefits. This kind of benefit is nearly impossible to measure. However, there are some proxies.

When looking at research as a contribution to education, numbers of degrees can be tallied and assertions about skills added to the Nation's work force can be made. When looking at research as creating new knowledge, one tangible "output" is papers published by scientific investigators to communicate new information to their scientific peers. Communicating the results of scientific research to colleagues through publication in the open literature is considered to be an important, if not essential, feature of good research practice.[34] Perhaps the best approach is to construct workable indicators and include a rigorous treatment of their uncertainties.

Bibliometrics

One tool that has been vigorously developed (especially in western Europe during the 1980s) for measuring the outcomes of research is bibliometrics, the statistical analysis of scientific publications and their attributes. Intrinsic to scientific publication is the referencing of earlier published work on which the current work is presumably based or has utilized in some way. References are a common feature of the scientific literature, and by counting how often publications are cited, bibliometrics can arrive at a weighted measure of publication output-not only whether publications have been produced, but also what impact those publications have had on the work of other scientists.[35]

OTA has explored several examples of new data sets that could be compiled using bibliometrics.[36] First, universities can be ranked according to an output or citation measure, the citation rates for papers authored by faculty and others associated with each institution. OTA drew on the large electronic database created and maintained by the Institute for Scientific Information (ISI).[37] Each institution in ISI's Science Indicators database, 1973 to 1988, was listed by its total number of cited

[32] For other suggestions, see Commission on Professionals in Science and Technology, op. cit. footnote 14. The above (hypothetical) data also raise the question of research outcomes--those relating to individual performance and that of other production units in the Federal research system.

[33] For a comprehensive review (now a decade old) of attempts at such measurement, see National Academy of Sciences, op. cit., footnote 8, especially ch. *2.*

[34] Robert K. Merton, "The Matthew *Effect* in *Science,* II: Cumulative Advantage and the Symbolism of Intellectual property," *Isis,* vol. 79, No. 299, 1988, pp. 606-623.

[35] Interpreting citation patterns remains a subject of contention. For caveats, see D.O. Edge, "Quantitative Measures of Communication in Science: A Critical Review,' History *of Science*, vol. 17, 1979, pp. 102-134; and S.E. Cozzens, "Taking the Measure of Science: A Reviewof Citation Theories," *ZSSK Newsletter, No. 7,* 1981, pp. 16-21. The definitive overview is contained in Eugene Garfield, *Citation Indexing: Its Theory and Application in Science, Technology and Humanities (New* York, NY: John Wiley & Sons, 1979). Also see Francis Narin, *Evaluative Bibliometrics: The Use of Citation Analysis in the Evaluation of Scientific Activity (Cherry* Hill, NJ: Computer Horizons, Inc., 1976).

[36] See Henry Small and David Pendlebury, "Federal Support of Leading Edge Research: Report on a Method for Identifying Innovative Areas of Scientific Research and Their Extent of Federal Support," OTA contractor report, February 1989; and Henry Small, "Bibliometrics of Basic Research," OTA contractor report, September 1990. For OTA contractor reports available through the National Technical Information Service, see app. F.

[37] The Institute for Scientific Information (ISI) database covers 7,500 journals published worldwide and indexes 800,000 new articles each year. The files derived from the ISI databases cover multiple disciplines and countries, and extend back to 1973. The analysis below is based on Small, op. cit., footnote 36. The Science Indicators File is a specially constructed multiyear file of publications from ISI's *Science Citation Index,* which contains a citation count time series for each paper in the file that has been cited one or more times for the years 1973 through 1988 inclusive.

Table 8-4-Mismatches in Rank Between Federal Funding and Average Citations: 1988

Of the top 100(107) federally funded universities, only 17 did not make the top 100 citation list. They are (with funding rank in parentheses):	Of the top 100 cited schools, only 17 did not make the top 100 (107) most funded schools. They are (with citation rank in parentheses):
Texas A&M (22)	University of California, SantaCruz(14)
University of Florida (45)	University of Oregon (23)
Woods Hole Oceanographic Institute (52)	SUNY Albany (43)
New Mexico State University (61)	Rice University (58)
Louisiana State University (72)	University of California, Riverside (59)
Utah State University (74)	St. Louis University (70)
North Carolina State at Raleigh (75)	Creighton University (80)
Virginia Polytechnic Institute and State University (76)	University of Notre Dame (81)
University of Kentucky (82)	University of Houston (82)
University of Dayton (86)	University of New Hampshire (84)
University of Nebraska at Lincoln (89)	University of Alaska (89)
Wake Forest University (95)	University of South Alabama (90)
University of Medicine and Dentistry of New Jersey (99)	College of William and Mary (93)
Washington State University (100)	Howard University (94)
University of Missouri, Columbia(101)	Brigham Young University (96)
Medical College of Wisconsin (104)	University of Delaware (97)
Rensselaer Polytechnic Institute (107)	University of Oklahoma (98)

NOTE: To compare the top 100 rankings, some institutions in the top 100 federally funded universities were disaggregate by campus of the State university system, e.g., the University of Texas, Austin. This added 7 entries to the top 100.

SOURCES: National Science Foundation, *Academic Science/Engineering: R&D Funds, Fiscal Year 1988*, NSF 89-326 (Washington, DC: 1990), table B-37; and Henry Small, "Bibliometrics of Basic Research," OTA contractor report, July 1990.

papers, the total citations received by all papers associated with each institution, and the ratio of number of citations to the number of publications, namely, the ***average citations per cited paper. This*** is a more discerning measure than either publication or citation counts alone. A ranking of institutions by average citation rates can be used in conjunction with the list of top universities in Federal R&D finding received to link inputs with outputs. (Appendix E lists the top 100 academic institutions ranked by their average citation impact for the period 1981 to 1988.) Table 8-4 lists the institutions, in 1988, that were among either the top 100 academic institutions in average citations or the top 100 receiving Federal R&D funds (again, see appendix B), *but not both.* Together, these measures illuminate differences in rank. The overlap in institutions suggests that the funding decisions by the Federal Government for the most part are leading to productive research. The mismatches may be indicative both of concentrated, rather than broad-based research productivity, and either some institutional "overachievement' or a substantial supplementation of Federal research support by State, corporate, and nonprofit sources.[38]

Trends in the average citation rate over time can also indicate how productive an institution has been in the published literature. The citation set can be analyzed by broad field or other variables to try to determine the cause of the changes (see box 8-A for profiles of four universities). Institutions can also be grouped to look at how, for example, "private institutions in the Southwest" or the national laboratories are performing as a category[39] (see figure 8-l). Many companies and other types of research organizations, despite proprietary inhibitions, also publish in the scientific literature and their work can be similarly aggregated and displayed (see figure 8-2). In another example for future exploration, programs receiving primarily directed funds or block grants (e.g., in agriculture) could be compared with those that are investigator-initiated. This comparison would help to test the claim that targeted appropriations (e.g., earmarking) lead to the production of inferior research.

38An institution that ranks high on funding and low on citation impact is not necessarily an underachiever. Some research is not readily published in the open literature, for proprietary or national security reasons.

39For example, the publication records and citation impact of National Aeronautics and Space Admini-stration research centers, 1973 to 1988, are examined in "NASA's Citation Impact Dims in 1980s, But Voyager Missions and JPL Shine," Science *Watch*, vol. 1, No. 9, October 1990, pp. 1-2, 7-8.

'loBiotechnology research is more often reported in the open literature than either research from electronics and computing firms or from Fortune 500 companies. Thus, the samples used in figure 8-2 may not represent the full range of research activity in these industries. Indeed, the most exciting results may be withheld from publication, but might be reflected in patents awarded later.

Box 8-A—Bibliometric Profiles of Four Research Institutions

OTA selected 19 institutions, based on historical patterns in their Federal funding profiles, to examine changes in research output and probe how they might be accounted for bibliometrically.[1] The institutions' publication and citation records were extracted to obtain a "citation impact" time series. This requires specifying four time points: a beginning and ending *cited* item period, and a beginning and ending *citing* item time period. This defines what items are eligible to receive citations and what journal publications are eligible to give them. OTA began with 1973, and defined the length of the period for analysis to be 8 years. This yields nine successive overlapping time windows that can be plotted as a time series or moving picture of the citation impact for each institution through 1988.

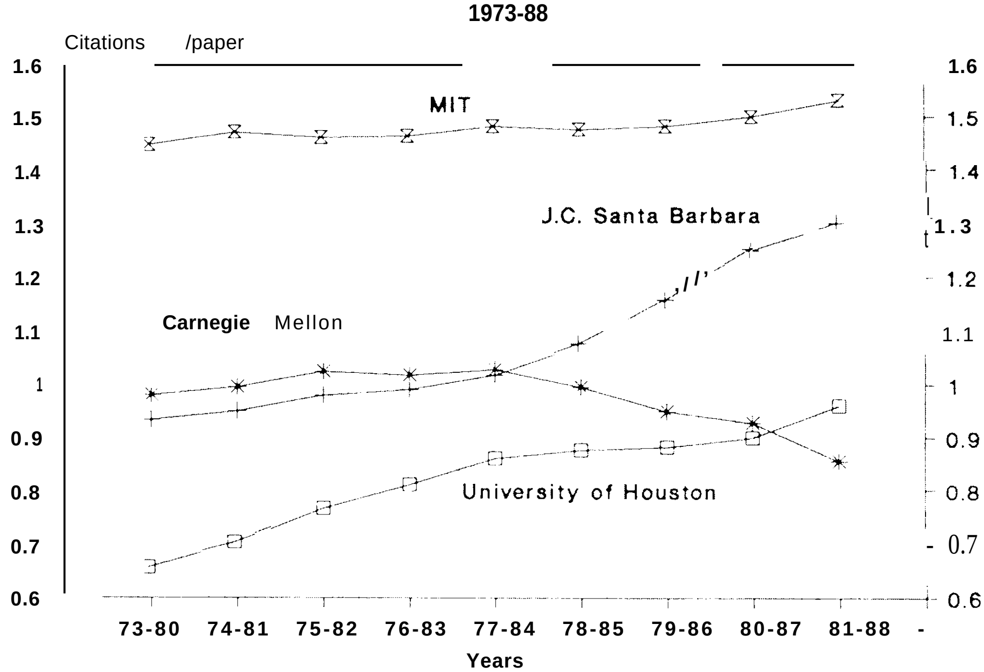
For example, counted in the first window were the number of cited papers published from 1973 through 1980 and the number of times those papers were cited by papers published in the same period. The ratio of these quantities is the mean citations per cited item for that time window. As a further normalization, each of the impacts is divided by the overall average for all U.S. papers for the specified time window, e.g., 1973 to 1980. The result is a measure of *relative impact.* Thus, a relative impact score of one signifies that the institution's average is identical to the average for all U.S. papers in the window. A score greater than one signifies an impact above the U.S. average, and a score below one an impact below the U.S. average.

The time series plots of relative impact for 4 of the 19 selected institutions are shown in figure 8A-1. To explain the trends observed in these graphs in terms of the fields of science involved, listings of the most cited papers were obtained for each institution, covering items cited 100 or more times, down to a maximum of 100 items.

1) *Massachusetts Institute of Technology (MIT)* has been consistently among the top 10 institutions for Federal R&D funds received. Like other top 10 institutions, which often produce relative impacts at the national average or above, MIT exhibits relative impacts in the 1.4 to 1.5 range. MIT also shows a modest gain in citation impact. Twenty-nine percent of its most cited papers are from the 1981 to 1988 period. Biomedicine has become stronger, while chemistry and geoscience have tapered off, and physics remained about the same.

[1] The following is based on Henry Small, "Bibliometrics of Basic Research," OTA contractor report, July 1990. Available through the National Technical Information Service, see app. F.

Figure 8A-1—Average Relative Citation Impact for Four Research Institutions, 1973-88



SOURCE: Henry Small, "Bibliometrics of Basic Research," OTA contractor report, July 1990.

2) *University* ***of Calofornia-Sarzta Barbara*** (UCSB) has improved its ranking among the top 100 recipients of Federal R&D dollars from 1%7 to 1984. OTA calls such institutions "upwardly mobile." Their patterns of research output are even more diverse than their relative gains in funding. UCSB displays a very marked increase in citation impact. It also has a very large number of 1981 to 1988 papers in its highly cited s@, 43 percent. Even more remarkable is the spread of these papers over various disciplines, with the emphasis on physics. Of the recent highly cited papers, 81 percent are in physics. Other areas represented include biomedicine, ecology, geoscience, and chemistry.

3) *Carnegie-Mellon University (CMU) has* been a top ***20*** recipient of Federal funds in engineering, and mathematics and computer science. However, it shows a decline in relative impact, beginning in the late 1970s. An analysis of the 78 papers cited 100 times or more shows that 15 percent of these papers are in the period 1981 to 1988. While 23 percent of the 1973 through 1980 papers were in the discipline of physics, only 8 percent are from physics in the later period. Chemistry, biomedicine, and computer science continue from earlier to later periods at comparable levels.

4) *University of Houston(UofH)* is a new comer to the select group of top 100 recipients of Federal R&D funds. It displays one of the most marked increases in citation impact of the institutions examined, although it started at a very low level. Its number of papers cited over 100 times is also small at 28. Nevertheless, 39 percent of these are from the recent period. Whereas chemistry and biomedicine were dominant early, physics (and more specifically, high-temperature superconductivity) account for most of the new highly cited papers (though biomedicine is also represented). Possibly a shift toward strengthening physics contributed to the increase m impact for this institution.

In some of these cases, it may be possible to attribute changes in citation impact to a shift in the field orientation of an institution. Such shifts maybe the result of deliberate organizational changes, or perhaps due to a resourceful faculty member who is able to move into new areas of research. One key to increasing impact is the ability to produce a continuing flow of **innovative papers that** influence researchers ''at the front. " This relates to the proportion of highly cited papers that are of recent origin. Reliance on aging 'classics" will not ensure an upward trend in impact. Another factor is field balance: some institutions seem to have strength across a number of fields, while other institutions focus on one or two seemingly to the exclusion of others. It is clearly more difficult for an institution to maintain excellence across a wide range of fields-the traditional mark of a *research* university--than to specialize in one or two.[2]

One lesson from the institutional profiles is that maintaining a high citation impact over a generation is difficult at best. The citation trends for UCSB and UofH confirm their upward mobility in research output as well as in Federal funding, in contrast to the citation trends at other institutions.

[2]Of the top 100 institutions in Federal R&D funding in 1988, only 39 had a relative impact score above the national average. Ibid.

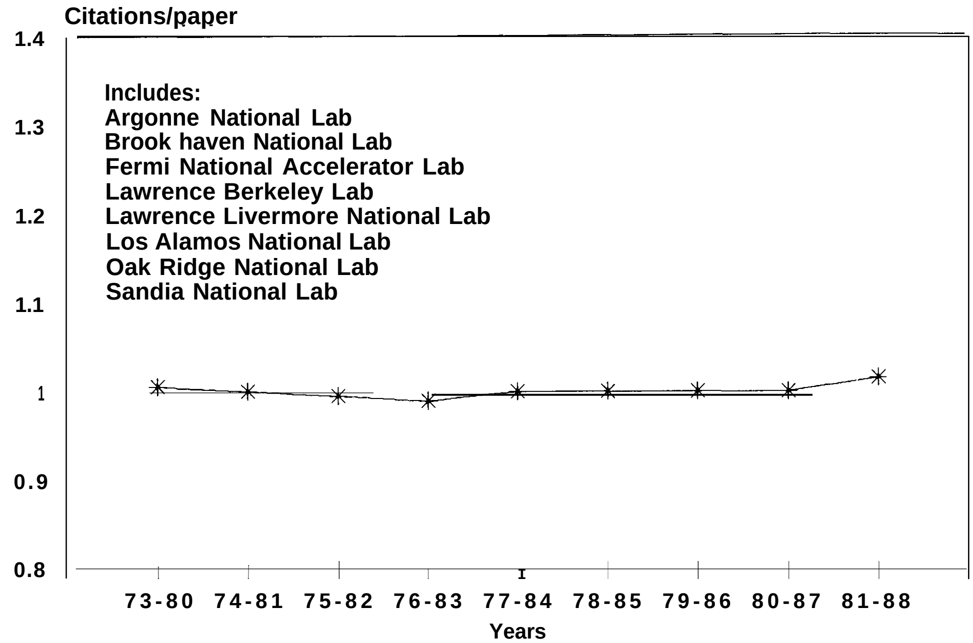
Not only can publishing entities be analyzed, but fields of study as well. For instance, ''hot fields,' in which the rate of publication and citation increases quickly over a short period of time, can be identified. Research areas such as high-temperature superconductivity emerge after a major discovery. Through co-citation analysis, papers can be sorted into ''clusters of publications that cite each other. These research clusters can be grouped further and mapped within disciplines.[41] In addition, related areas that contribute to the work can be identified and linked across disciplinary boundaries. For example, high-temperature superconductivity research has been connected with work in ceramics, thin films, polymers, and other diverse areas.[42]

If the papers comprising a cluster cite their sources of funding, an estimate can be made of Federal support of the research represented by the clusters. To demonstrate this method, OTA requested that a small sample of papers published be searched for funding information in a cluster repre-

[41]For elaborations of the algorithm and the interpretation of resulting co-citation maps, see Henry Small and B.C. Griffith, 'The Structure of Scientific Literatures I: Identifying and Graphing Specialties," *Science Studies,* vol. 4, 1974, pp. 17-40; and Henry Small and Eugene Garfield, "The Geography of Science: Disciplinary and National Mappings," *Journal of Information Science, vol.* 11, December 1985, pp. 147-159.

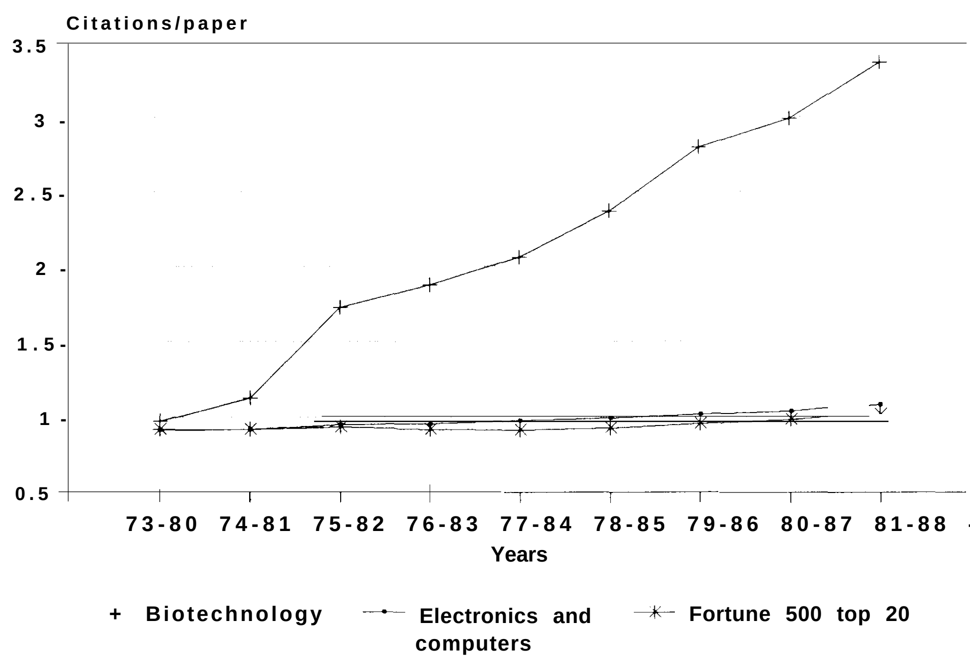
[42]Small and Pendlebury, op. cit., footnote 36. These connections have been confirmed independently through analysis of other, nonbibliometric data. See John M. Rowell, ' 'Superconductivity Research: A Different View, ' *Physics Today*, November 1988, pp. 38-46; and Dorothy Robyn et al., ''Bringing Superconductivity to Market,'' *Issues in Science& Technology*, vol. 5, No. 2, winter 1988-89, pp. 38-45.

Figure 8-l—Relative Citation Impact for National Laboratories



SOURCE: Henry Small, "Bibliometrics of Basic Research," OTA contractor report, July 1990.

Figure 8-2—Relative Citation Impact for Three Industries



NOTE: *Biotechnology companies include:* Amgen, Biogen, Biotech Labs, California Biotech, Centocor, Cetus, Chiron, DNAX, Genentech, Genetics Institute, Immunex, and Molecular Genetics Inc. Electronics *and computer companies include:* IBM, Digital Equipment, General Electric, Westinghouse Electric, Eastman Kodak, and Xerox. *Fortune 500 Top 20 include:* General Motors, Ford Motor, Exxon, IBM, General Electric, Mobil, Philip Morris, Chrysler, DuPont de Nemours, Texaco, Chevron, Amoco, Shell Oil, Proctor & Gamble, Boeing, Occidental Petroleum, United Technologies, Eastman Kodak, USX, and Dow Chemical.

SOURCE: Henry Small, "Bibliometrics of Basic Research," OTA contractor report, July 1990.

Photo credit: U.S. Department of Energy

Researcher holds a piece of superconducting tape. Scientists must be able to make ceramic superconductors in a variety of forms to be useful—from thin films for electronics to casts for accelerator cavities. The development of ceramic superconductors has been an outcome of superconductivity research.

senting research directly related to high-temperature superconductivity. (Similar analyses were conducted in four other research areas.[43]) Roughly one-half of the most cited papers in 1985 to 1987 (77 of 139 papers) were coded for funding information and a random sample of the papers that cited them in 1989 were included (95 of 1561). More than one-half of the funding acknowledgments were to Federal funding agencies (with over one-half to NSF, slightly under one-third to the Department of Energy (DOE), and significant contributions from the Office of Naval Research and the National Aeronautics and Space Administration). Corporations, primarily IBM and AT&T, funded another one-third of the papers; Federal laboratories, private foundations, and foreign sources supplied the remaining funds. The Federal Government is a continuing catalyst of high-temperature superconductivity research.

The degree to which a field is international in effort can also be indicated through the nationality of authors. Again, for the high-temperature superconductivity cluster, the most cited papers in 1985 to 1987 were from the United States (64 percent), followed by France (8 percent), Japan (8 percent), Switzerland (4 percent), Canada (4 percent), and the United Kingdom (3 percent). The institutions in which these papers most often originated were AT&T (12 percent), IBM (12 percent), University of Houston (5 percent), University of Tokyo (4 percent), Bell Communications (4 percent), and the University of California-San Diego (3 percent). Countries citing the papers were more diverse, with the United States at 43 percent; Japan, France, and the Federal Republic of Germany at 5 to 6 percent each; the U.S.S.R. and India at near 5 percent; and the Peoples Republic of China at 4 percent. Similar diversity is seen in the institutions where these papers originated.

With these types of analyses, bibliometrics could perhaps be used to track the evolution of fields and subfields-by research topic and national or institutional authorship. However, there are significant disadvantages to bibliometrics, which also must be recognized.[44] In particular, citations are not made in a uniform way in the scientific community, and neither is allocation of authorship. Also, the same discovery may be cited in different ways in different publications. Consequently, only in the aggregate and when comparing similar fields with similar citation practices can judgments of hot fields, influential papers, and prolific authors be made with confidence.[45] The utility of bibliometrics should be seen as ''value-added' to policy analysis, not as stand-alone information.

Other Measurement Techniques

Another genre of outcome measures focuses on the research-technology interface. There are many examples of data that could be collected to illuminate the relationship of research to other parts of the development cycle. Complicating features, however, include technological choice within private or public firms that develop technology, utilization of science and engineering talent, and the transfer of knowl-

[43] Small and Pendlebury, Op. cit., footnote 36.

[44] As Eugene Garfield, founder of the Institute for Scientific Information (which pioneered citation databases and their analysis), warns: ''You can misuse citation analysis easily—that's the story of my life. ' Quoted in Gina Kolata, "Who's No. 1 in Science? Footnotes Say U.S., " *The New York Times,* Feb. 12, 1991, pp. Cl, C9.

[45] See Susan E.Cozzens, ''Literature-Based Data in Research Evaluation: A Manager's Guide to Bibliometrics,'' final report to the National Science Foundation, Sept. 18, 1989.

edge from research centers to other sectors of the economy.[46]

Sponsors, at least for basic research, have little control over the execution of the projects they support. So lack of payoff may be unrelated to the intrinsic merits of project design and substance and have more to do with the differential competence and efficiency of performers. But no sponsor can ascertain the most competent, creative, and efficient of performers.[47]

In the case of public programs with firm measures of outcomes, negative evaluations suggest termination. But for programs whose output is information, the situation is highly problematic. Information volume and quality might be low, but this may be because the overall level of resources is too low. Or a program may have technical inefficiency due to poor management. Or the *lack* of results may itself have high scientific or technological value. Since research deposits knowledge into the scientific literature, it may take years to be applied to other problems. Some ideas are premature, and others remain invisible to specialists in fields different from the authors' own. Recognition of the utility of research-both intended and unintended-is often delayed.[48] This does not depreciate its value, but does impede its use.

Historically, science and technology are full of sudden reversals about the value of information produced by past research. Testing hundreds of compounds for superconductivity was, until recently, not considered high-grade science, but mundane science. And any evaluation of this work would have suggested that this kind of research was not worth much investment. Similarly, the funding of the early recombinant DNA projects was not done in the name of expected high payoffs. Certainly no one at the time imagined a biotechnology industry as the result. Thus, the ability of research evaluations to provide credible estimates of the incremental information gains from additional funding is weak. Federal agencies tend to use an insurance principle and spread resources widely to ensure that no reasonable bets are overlooked. (From one perspective, this is risk-averse; from another it is risk-taking, because ideas from out of the mainstream can be supported.)

Bibliometrics and production function data on the research-technology interface are examples of tools that could be used to evaluate outcomes. While not exhaustive, they illuminate different aspects of science as a process and the utility of research performance.[49] As with the examples discussed above, data collection can be improved when the user of the data and the purpose are targeted. The next section explores how data on the Federal research system is employed by policymakers and how new data could aid the transition from analysis to decisionmaking.

Utilizing Data

In a policy context, information must be presented to those who are in positions to effect change by allocating or redirecting resources.[50] In the diverse structure of the Federal research system, many actors play roles in research decisionmaking at many different levels. These actors require data reported in various forms and units to make decisions. For example, an agency program manager requires data specific to the purview of his or her programs, while OMB and OSTP must be aware of trends in science

[46]Given the large, but cheap increases in computing power, various models are commonly used by management analysts for deciding on ex ante investments, but these techniques remain very sensitive to subjective and highly uncertain estimates of technical and market success. One notable exception is Edwin Mansfield, "The Social Rate of Return From Academic Research,'' *Research Policy,* forthcoming 1991.

[47]Recent advances in methods of measuring returns to basic research have centered on sophisticated econometric techniques for estimating *production functions (e.g.,* measures of the economic impact of research). Since the marginal value of research is heavily dependent on downstream events, production functions could be embedded in fuller models of information flow and economic behavior. In addition, literally hundreds of *quantitative project selection* methods exist in industry for guiding investments. Methods include elaborate goal progamming and analytical hierarchy models; the techniques are often known as return-on-investment, impact matrices, or checklists. See Harvey Averch, "Policy Uses of 'Evaluation of Research' Literature," OTA contractor report, August 1990. Available through the National Technical Information Service, see app. F.

[48]See Gunther S. Stent, ''pre~~ty and Uniqueness in Scientific Discovery," *Scientific American,* vol. *227,* December 1972, pp. 84-93; and Julius H. Comroe, "The Road From Research to New Diagnosis and Therapy," *Science,* vol. 200, May 26, 1978, pp. 931-937.

[49]The application of bibliometrics t. patenting behavior, i.e., measuring the dependence of patents on the scientific literature, has pioneered new ways of thinking about the diffusion and application of research knowledge. See Francis Narin et al., "Patents as Indicators of Corporate Technological Strength,'' *Research Policy, vol. 16, 1987,* pp. 143-155; and Zvi Griliches, ''Patent Statistics as Economic Indicators: A Survey,' *Journal of Economic Literature, vol. 28, No. 4,* December 1990, pp. 1661-1707.

[50]For example, see Carol H. Weiss, "Improving the Linkage Between Social Research and Public Policy,' *Knowledge and Policy: The Uncertain Connection,* L.E. Lynn (cd.) (Washington DC: National Academy of Sciences, 1978), pp. 23-81.

that span broad fields, institutions, and agencies, as well as those that apply only to specific fields, performers, and sponsors. Timely data are similarly important. For instance, world events can alter the perception and utility of even the best information and analysis (see box 8-B).

Providing data at each of these levels is a large task, but one that is essential. As seen with projected shortages of scientific and engineering personnel, trends are often specific to disciplines and to types of institutions, and decisions that take into account these differences would best address impending problems. Enhanced internal agency data collection would help to disaggregate and distinguish trends most relevant to the agency.

As well as targeting data collection to the needs of decisionmakers, the data above must address policy-relevant questions, i.e., be used evaluatively, as well as illuminate significant trends.[51] Thus, there has developed a distinction between standard data collection (i.e., tabulations on one variable, such as Ph.D.s awarded) and the development of *indicators*-data presented in such a way (e.g., comparisons between variables) as to suggest patterns not otherwise discernible. For instance, data on the rising cost of equipment in a specific field (or the rate of change in this cost) have little meaning unless compared with the cost (and percent change) of equipment in other fields. A measure of the ***relative*** cost of equipment in different fields would indicate the need to make special provisions for equipment in select fields. Similarly, data on the decline of baccalaureate degrees in a natural science field are more useful when they are compared to other broad fields, and judged in terms of absolute and relative declines and stability.

Indicators do not necessarily prescribe a course of action, but they warn of possibly significant trends. As part of the decision process, they offer "usable knowledge."[52] At present, indicators on the Federal research system are neither comprehensive nor objective-driven.[53] The focus of the *Science & Engineering Indicators* volumes has been less on indicators than on data. Indeed, SEI is a statistical reference book that collates available data on the research system. Additional efforts to produce indicators, especially on research performers, could greatly enhance utilization and action by decision-makers.

New Indicators

NSF, specifically the Special Data Group attached to the Director's Office, has recently attempted to develop new indicators related to research participation at NSF during the 1980 to 1989 decade.[54] These indicators are defined and summarized in table 8-5. Though their meaning is not always straightforward,[55] these indicators represent a significant advance in reconstructing trends in NSF proposal and award activity.

The first indicator in table 8-5, the Proposal Success Rate, is driven by the change in the number of proposals submitted. At NSF, this number increased by 30 percent during the decade. Over 20 percent of those originally declined resubmit proposals to NSF (with an equal proportion submitting elsewhere). While the Proposal Success Rate declined from 38 percent in the beginning of the decade to 31 percent in 1989, PI (Principal Investigator) Success Rate from 1980-82 to 1987-89 remained above 40 percent. The PI Success Rate indicator allowed NSF to conclude that more PIs are being funded, but they facer stiffer competition to win awards. However, the relation between these two Success Rate indica-

---

[51] The methodological pitfalls in applying data to evaluate national or institutional research performance are illustrated in John Irvine et al., "Investing in the Future: How Much Governments Pay for Academic Research," *Physics Today*, September 1990, pp. 31-38; and Jeremy Cherfas, "University Restructuring Based on False Premise?" *Science*, vol. 247, Jan. 19, 1990, p. *278.* For further discussion, see David C. Hoaglin et al., *Data for Decisions: Information Strategies for Policymakers* (Cambridge, MA: Abt Books, 1982).

[52] Charles E. Lindblom and David K. Cohen, Usable Know/edge: *Social Science and Social Problem Solving* (New Haven, CT: Yale University Press, *1979).*

[53] Cozzens, op. cit., footnote *4,* pp. 15-17.

[54] The Special Data Group is part of the Comptroller's office at the National Science Foundation. It works independently Of two Other staffs in the Scientific, Technological, and International Affairs Directorate that also develop science indicators—the Science Resources Studies Division (home of the *Science & Engineering Indicators volumes)* and the Policy Research and Analysis Division (which in 1990 issued the data-laden, *The State of Academic Science and Engineering).*

[55] These are based *on* National Science Foundation, "NSF Vital Signs: Trends in Research Support, Fiscal Years 1980-89," draft report, Nov. 13, 1990.

Box 8-B-War as a Wild Card: The Impact of the Persian Gulf Conflict on Science and Technology

After World War II came *Science--The Endless* Frontier.[1] A nation grateful for its success, and newly aware of its responsibilities in the world, decided that science was an important part of that world, and that science would benefit from government funding. Despite the negative impact of Hiroshima and Nagasaki, science came out of the war with a positive image, a great deal of momentum, and a strong basis for Federal support.

Vietnam, an unpopular war with unclear objectives, came with defoliation, napalm, Agent Orange, and accusations of environmental degradation. Science and technology, were cast in a negative light of the atomic bomb drops on Japan and tarred with the brush of destruction. Antiscience became part of the antiwar movement, and the remnants of this antiscience sentiment are still with us today.

The United States has just waged the most technological war the world has yet known.[2] For exarnple, after years of controversy and failed test results, the Patriot missile served a cogent strategic and political purpose. Even with its outdated technology, the Patriot strengthened the claims of some that electronic warfare has come into its own. What such success may mean for the future image of science and technology is as yet unknown.

War is a wild card. Its effects on the populace at large (and on potential science and engineering students in particular) are difficult to predict. War is a reminder that events outside of science can reverberate in many ways--changing images and attitudes-for a long time to come. Analyses and reports on science and technology, such as this one, can only begin to measure, much less anticipate, these impacts.

[1] This box is based on a draft report by Alan McGowan, president, Scientists' Institute for Public Information, and a member of the Advisory Panel for this OTA report.

[2] See Sidney Perkowitz, "The War Science Waged," *The Washington Post*, Mar. 3, 1991, p. C2.

tors and inferences about PI proposal-writing behavior is unclear for decisionmaking.[56]

The Continuity of Support indicator shows that nearly one in three of the PIs with NSF support in 1980 were still receiving support in 1989. The Flexibility of Support and Continuity of Support indicators together measure the balance between providing stable support to (established) investigators and retaining the ability to bring new investigators into the NSF funding system. Funding of new PIs fluctuated with the decline or growth in NSF obligations. Also, directorates with higher success rates (Geosciences, and Mathematics and Physical Sciences) ranked lower in Flexibility, because they are supporting a more established group of researchers.

The Award Size/Duration indicator reflects how NSF responded to increased demand for funding. Early in the decade, the number of awards was held constant but the award amounts were increased; later more proposals were funded and median award size did not grow. Throughout the decade, median annual award amount represented 80 to 85 percent of the *requested* amount.[57]

Indicators are best used to *monitor* trends, especially if they could be extended to other agencies as well.[58] This would help to complete the picture of PI proposal-writing strategy and the distribution of

[56] Linda Parker, Comptroller's Office, National Science Foundation, personal communication, January 1991, suggests that more proposals are being submitted and the principal investigator population is increasing, but resubmissions (of previously declined proposals) account for only 20 percent of the growth. James McCullough, Comptroller's Office, National Science Foundation personal communication, March 1991, reports that 30 percent of the proposals received by the National Science Foundation (NSF) in any year came from researchers who had not submitted in the previous 5 years (which NSF defines as 'new investigators' and another 20 percent are received from researchers who submitted only one proposal. The supply of "new blood" and demand for funding seem hearty.

[57] Award amounts are negotiated. Principal investigators inflate their requests in the expectation that they will not receive "full" funding. If declined their resubmissions tend to feature smaller budgets. In multiyear (e.g., 2 to 3 year) awards, which are now typical at the National Science Foundation annual project budgets are fixed at the outset of the award, subject only to across-the-board cuts in succeeding years. (This contrasts with the National Institutes of Health's practice of annual downward negotiation in multiyear awards.) Robert P. Abel, Office of Budget and Control, National Science Foundation personal communication, July 1990.

[58] The National Science Foundation cautions about the interpretation of indicator trends. Changes maybe due to: a) an across-the-board budgetary upheaval, e.g., the Gramm-Rudman sequester of 1986, which reduces the capacity to fund; b) a targeted increase or decrease in appropriations to a directorate (or more generally, any agency line item) or c) agency reorganization or creation of programs tbat shifts proposals and awards in ways that affect disaggregate uses of an indicator.

Table 8-5-New Indicators of Research Activity at NSF: Fiscal Years 1980-89

Indicator	Definition	Comment
Proposal Success Rate	Ratio of awards to total actions (new award and decline decisions) on competitive (peer- or merit-reviewed) proposals	Measures at an aggregate level proposal activity and awards that result in National Science Foundation (NSF) commitment of new funding. Interpretation of the indicator is not straightforward. Assumes estimates of growth in the research work force, rising costs of research, change in proposal review criteria and a proliferation of special award categories (set-asides). The indicator does require knowledge of agency context.
PI Success Rate	Number of principal investigators (PIs) who are successful (within a 3-year period) in winning an award divided by total number of investigators submitting proposals (within the same 3 fiscal years)	Contrasts with Proposal Success Rate, which indicates NSF action generated by proposals submitted to it. Measures effort and success of the research population to gain NSF support, including changes in mean number of submissions needed to win one award.
flexibility: New PI Funding	Percentage share of total award dollars going to PIs who have not had NSF support in the previous 5 years	Indicator is most revealing when compared to other indicators. Definition of "New PI" is only a proxy for "young investigators."
Continuity of Support	Percent of principal investigators receiving support at the start of a time period who are still receiving support at the end of the period	Indicator complements "flexibility," which measures awards to investigators without prior awards. It can be indexed to any cohort of grantees and calculated for prior or succeeding years. Indicator identifies investigators with sustained support.
Award Size/ Duration	Total award dollars divided by total award years (duration)	Change from total dollars obligated in a particular fiscal year to amount of award over its lifetime provides a more accurate picture of support as experienced by the PI. Award size and duration affect the character and pace of research activity. Reduced award size may affect number of proposals written, while reduced duration may affect frequency of proposal writing. Both require investigator time for research. This indicator assumes no other, i.e., non-NSF, source of research support. It requires caution in making inferences about time spent in proposal writing.

SOURCE: Office of Technology Assessment, 1991, based on National Science Foundation, "NSF Vital Signs: Trends in Research Support, Fiscal Years 1980-89," draft report, Nov. 13, 1990.

research demand by field and agency. Thus, indicators could become an important part of the priority-setting process. Perhaps this argues for OSTP to coordinate across agencies the development and presentation of a prescribed set of indicators. Disaggregate to reflect disparate agency structures, such as directorates and divisions at NSF, such indicators could also help portray variations in fields and research communities. Sensitivity to such disaggregations maybe most instructive for research funding policy. As an NSF task force recently put it:

> Part of the problem is a lack of understanding of the actual size of the research community and what fraction of a specific community should be funded. A clear, coherent picture of community size is essential. How many grants should be awarded and at what budgets? Should NSF fund all fields or make choices predicated on the investments of others? This "snapshot" of the community should be updated regularly in order to indicate achievement or changes that might be necessary to minimize confusion with respect to overall NSF policy issues.[59]

OTA concurs. Such baseline information should be routinely available to decisionmakers in the 1990s. Overall, *sets* of indicators that draw on these data are preferable to single measures. With this in mind, OTA (building on the new NSF indicators reviewed above) suggests the following four sets. They could be compiled and analyzed by all of the research agencies or by OSTP, which would complement existing indicators constructed and reported in

[59]Some of the indicators presented above were indeed used for an inhouse evaluation of how to streamline the workload of the National Science Foundation' sprogram staff and the external research community. Short-term recommendations ecus on simplifying the proposal preparation and review process, including budgets; long-term recommendation include ways of balancing support modes and restructuring grant types (size and duration), especially cross-directorate programs. See National Science Foundation, *Report of the Merit Review Task Force,* NSF 90-113 (Washington, DC: Aug. 23, 1990).

292-863 0 - 91 - 9

SEI, to provide windows on various segments of the Federal research system:

- "Active research community" indicators, which would estimate the number of researchers actively engaged in federally funded research (e.g., PIs currently supported by one or more Federal grants plus those with a research proposal pending at a Federal agency, and proportion of research time that is federally funded).
- "Research expenditure" indicators to recalibrate Federal expenditures by line item of research budgets (e.g., salaries, equipment, and facilities) and by broad field.
- Federal "proposal pressure" indicators, e.g., proposals submitted to the Federal Government per investigator, ratio of Federal to (investigator's self-reported) non-Federal proposals and projects in force at the time of submission, and fraction of requested project budgets actually awarded by the funding agency.
- "Production unit" indicators, e.g., the size of the research team or other performing unit supported in part through Federal grants (disaggregate by subfield, institution type, and agency source).

The combination of such indicators would estimate more precisely the changing parameters of the Federal research system.[60] This information could be invaluable to policymakers concerned about the health of certain sectors of the system. To produce such information, as part of ongoing agency data collection and NSF responsibilities for collation and presentation, extra resources would be needed. They might come from streamlining current NSF data and analysis activities, such as a reduction in the number of nonmandated reports issued annually, or designating a special unit, much like the Science Indicators Unit, to expand its inhouse and extramural "research on research. ' If there is a premium on timely information for research decisionmaking, it must be declared (and funded as) a Federal priority.

The utility of data is judged by many participants in the system: the needs of Congress are usually agency- and budget-specific;[61] the agencies, in contrast, worry about the performance of various programs and their constituent research projects. Data converge in one other underutilized source of information-the evaluation of research projects and programs after they have (or have not) produced results.

Evaluation of Research

While data and indicators can provide valuable information on aggregate trends in the research system, it is much more difficult to evaluate specific research investments in agency programs or projects (apart from charges of fraud, incompetence, or other gross flaws, which are investigated as part of the congressional oversight function). The returns from the performance of research to society are quite diverse. They include economic, health, security, educational, and many other benefits. Because research is a public good, there is little incentive for private investment (in terms of social returns). In 1986, OTA looked at ways to measure the returns from public investments in research:

> In summary, OTA finds that. . . the factors that need to be taken into account in research planning, budgeting, resource allocation, and evaluation are too complex and subjective; the payoffs too diverse and incommensurable; and the institutional barriers " too formidable to allow quantitative models to take the place of mature, informed judgment.[62]

Five years have passed since OTA announced this conclusion. However, demand for research evaluations has increased in all countries that make significant investments in research. The reasons for increased demand are the same: budgetary constraints, greater accountability to sponsors, and the

[60] For example, what would be the indications that growth in research productivity is slowing or that the size Of a research community is precariously large or small relative to the resources for supporting it? See Colleen Cordes, "Policy Experts Ask a Heretical Question: Has Academic Science Grown Too Big?" *The Chronicle of Higher Education, vol. 37, No. 2, Nov. 21,* 1990, pp. Al, A22.

[61] As several National Science Foundation staff have indicated to OTA project staff (personal communications, October-December 1990), the President's Science Advisor draws heavily on unpublished and newly published Science & *Engineering Indicators* (SEI) data in preparing and presenting the Administration's policy proposals at congressional "posture hearings" early in the annual authorization process. Indeed, the production cycle of SEI is geared to delivery of the volume as an input to this budget process.

[62] U.S. Congress, Office of Technology Assessment, *Research Funding as an Investment: Can We Measure the Returns?* OTA-TM-SET-36 (Washington DC: U.S. Government Printing Office, April 1986), p. 9.

Photo credit: U. S. Department of Agriculture

Oversized thornless blackberries, tiny strawberry plants (in jars), and star-shaped slices of carambola (a tropical fruit now grown in Florida) are examples of outcomes of the Department of Agriculture's Agricultural Research Service programs.

desire for increased rationality in decisionmaking.[63] In response, many funding agencies—here and abroad—have formed evaluation units.[64] Research evaluators and designers of science indicators have carried out substantial work on measuring scientific and technological performance. There have been refinements in existing methods for evaluating research impacts ex post. Nevertheless, examination of the published and unpublished literature on research evaluation methods between 1985 and 1990 suggests that OTA's conclusion still stands: evaluation methods are not cited as guides to research decisionmaking by national governments.[65]

Since 1985, no methods have been invented that more definitively measure the scientific or social value of past research investments. By ''definitively measure, OTA means that evaluation outcomes, whether positive or negative: 1) will be accepted without lengthy technical and political disputes among sponsors, clients, and constituents, and 2) will provide unambiguous direction in resource allocation or other kinds of decisions. While computer modeling permits greater use of ex ante (i.e., before the research project is attempted) project selection methods and ex post evaluation methods, the evidence is sparse that there is much short-term payoff to public or private sector research administrators from making greater use of them.[66]

The problem, however, may reside more with decisionmakers than the evaluation tools (and results) at their disposal. Research administrators have little incentive to use current evaluation technologies for making decisions about awards or level of project allocations. This lack of incentive persists because research evaluation ". . . occurs in a political context; is inevitably seen as *post hoc* justification for decisions (unrelated to the content of the evaluation); and should be anticipatory, designed to answer specific questions raised by superiors within the organization as well as critics from outside."[67]

Below, OTA first describes evaluation practices and processes around the world; then considers incremental improvements in research evaluation methods since 1985; and finally suggests that research evaluation faces certain inherent limits. These limits make it unlikely that however precise the measurement of average or incremental monetary or informational returns, they may not be embraced. Nevertheless, rather than a means of computing returns on past public investments or guiding prospective ones, research evaluation may help Federal funding agencies keep 'on their toes,' just as environmental impact statements impel

---

63 Even countries that do not make significant research investments claim they evaluate. For example, Finland, Hungary, and Poland reported on their evaluation efforts in the ''ECE Seminar on Evaluation in the Management of R and D, Apr. 3-7, 1989, " unpublished proceedings.

64 For discussions, see Ciba Foundation, *The Evaluation of Scientific Research* (New York, NY: John Wiley& Sons, 1989); and M. Gibbons and L. Georghiou, *Evaluation of Research: A Selection of Current Practices* (Paris, France: Organisation for Economic Cooperation and Development, 1987).

65 An annotated bibliography on evaluation of research, 1985 to 1990, contains 147 works published in scientific journals and government documents. It is the basis for this section. See Averch, op. cit., footnote 47.

66 K.M. Watts and J.C. Higgins, ''The Use of Advanced Management Techniques in R&D,'' *Omega*, vol. 15, January 1987, pp. 21-29. This survey, consistent with past surveys, shows that R&D administrators prefer simple, transparent methods of project selection. Interestingly, OTA consistently finds in discussions with policymakers in the executive and legislative branches that the proposition "high yield from Federal investment in R&D'' is taken as axiomatic. The issue is not *whether* to fund, but *what* and *how.*

67 Chubin, op. *Cit.*, footnote 18, p. 84. Also see Harold Orlans, "Neutrality and Advocacy in Policy Research,'' *Policy Sciences*, vol. *6, 1975,* pp. 107-119.

agencies to assess the effects of their programs on the environment.[68]

Evaluation in Other Countries

Table 8-6 summarizes the characteristics of the evaluation process among major scientific and technological powers.[69] Overall, in countries with parliamentary governments, national priority setting in research becomes a tool both of project selection and research evaluation.[70] The United Kingdom and France present contrasts in approaches to evaluation—the former contracting for outside analysis, the latter incorporating analysis of research outcomes into the government's apparatus and process for policymaking. Smaller countries, such as The Netherlands and Sweden, must be selective in the areas of research they target. If a' 'critical mass' of researchers is not available, collaboration in cooperative international projects becomes the only outlet for research participation.[71]

U.S. researchers have historically operated at the frontiers of knowledge, and other countries have adjusted their own research ventures as scientists in the United States and other scientifically advanced nations uncover promising areas. "Thus it is easier for countries off the frontier to identify what they want to pursue. Of course, the U.S. enterprise is so large that it uncovers more areas than smaller countries can afford. So they have a much more difficult choice than the United States in determining exactly what to pursue. '[72] Now that the U.S. role as a research performer is changing in some areas so that U.S. scientists may not always be at the forefront,[73] the time maybe ripe to review the place of research evaluations-especially relative to advances in other countries-in agency decisionmaking.

An Approach To Evaluating Basic Research Projects

Because of uncertainties attached to each and every research investment, procedures for their evaluation can be augmented by using ex post review by peer researchers and citation evidence jointly.[74] One approach would apply the following seven criteria weighted by the priority assigned to each:

1. value of the information produced: salience, relevance, importance-of both positive and negative results-to the field;
2. probability of use;
3. originality of results;
4. efficiency and cost;
5. impacts on education and human resources;
6. impacts on infrastructure and capability to carry out additional research in the future; and
7. overall scientific merit.

The overall ex post peer evaluation of particular projects can be compared with associated bibliometric information.[75] If this comparison indicates the same quality for a project, then the sponsor can have

[68] See, for ~-pie, S. Taylor, *Making Bureaucracies Think: The Environmental Impact Strategy of Administrative Reform* (Stanford, CA: Stanford University Press, 1984); E.N. Goldenberg, "The Three Faces of Evaluation" *Journal of Policy Analysis and Management,* vol. 2, summer 1983, pp. 515-525; and R.V. *Bartlett, Policy Through ImpactAssessment: InstitutionalizedAnalysis as a Policy Strategy* (New York, NY: Greenwood Press, 1989). The idea is to inducedecisionmakers to incorporate research evaluation information into their planning, not to impose the information under threat of punishment.

[69] The match between countries discussed in app. D of this report and those profiled in table 8-6 is not perfect. For recent comparative analyses among some of the countries considered here, see Department of Trade and Industry, *Evaluation of R&D—A Policymaker's Perspective* (London, England: Her Majesty's Stationery Office, 1988); L.L. Lederman et. al., "Research Policies and Strategies in Six Countries: A Comparative Analysis," *Science and Public Policy,* vol. 13, No. 2, April 1986, pp. 67-76; B.R. Martin and J. Irvine, *Research Foresight: Creating the Future* (London, England: Frances Pinter, 1989); and A.F.J. van Raan (cd.), *Handbook of Quantitative Studies of Science and Technology* (Amsterdam, The Netherlands: North-Holland, 1988).

[70] Averch writes: "Most European countries have ministries of science and technology that control the flow of resources for S&T. These ministries usually construct S&T plans correlated with economic plans. They are far more able to direct research programs at universities and industrial laboratories." See H.A. Averch, "New Foundations for Science and Technology Policy Analysis," paper presented at the Conference on The Mutual Relevance of Science Studies and Science Policy, Blacksburg, VA, May 12, 1989.

[71] See, for example, Jean-Francois Miguel, "Indicators to Measure Internationalization of Science," unpublished paper, 1989; aud Francis Narin and Edith S. Whitlow, *Measurement of Scientific Cooperation and Coauthorship in CEC-Related Areas of Science,* vol. 1 (Luxembourg: Commission of the European Communities, May 1990).

[72] Averch, op. cit., footnote 70, p. 15.

[73] Some examples of hot fields dominated by non-U.S. researchers are presented in Small, op. cit., footnote 36.

[74] See, for example, R.N. Kostoff, "Evaluation of Proposed and Existing Accelerated Research Programs of the Office of Naval Research,' *ZEEE Transactions on Engineering Management,* vol. 35, November 1988, pp. 271-279.

[75] See John Irvine and Ben Martin, *Foresight in Science: Picking the Winners* (London, England: Frances Pinter, 1984).

Table 8-6-Characteristics of the Research Evaluation Process for Selected Countries

Government	Methods of evaluation	Types of research evaluated	Reported utility	Central evaluation units	Government standards	Reporting
United Kingdom. . .	Peer review citation; publication; rate-of-return; patents; check-lists; market outcomes	Large projects; programs; universities; laboratories	Improve policy decisions	Assessment Office within Cabinet office; Department of Trade and Industry has central unit; some departments have Chief Scientists and Departmental Review Committees	Definitions of good practice	Some public
FRG	Economic and market indicators; ex post peer review for basic research projects; special committees; evaluations of disciplines; bibliometrics; patents; market outcomes	Large and small projects and programs	Improve policy decisions	Federal Ministry of Service and Technology (BMFT) has central unit	No	Some public
Japan.	Consistency with plans developed by "foresight"; market tests for applied commercial projects	Projects; priority programs	Planning	New central science policy unit has some evaluation responsibilities	No	Some public
Netherlands	Peer review; publication-citation indicators for basic research; client satisfaction or utility for applied projects; profits earned from research contracts	Projects; basic research programs, universities,industrial research	Ensure consistency with plans; assist with allocation decisions	Yes in public agencies and universities	No	All public

SOURCE: Harvey Averch, "PolicyUses of 'Evaluation of Research' Literature," OTA contractor report, August 1990. Note that most of the literature on Japan discusses their R&D planning processes and the use of "foresight" methods. There is a less open literature on ex post evaluation.

greater confidence that the project is, in fact, of that quality. If the two measures are not congruent, then the project can be subjected to more intense analysis to explain the discrepancy .76 In addition, an agency program can submit information on funded projects to experts working at the research frontiers to see whether its structure and content are judged as significant contributions to the field. This is expensive, but has been attempted, for example, for DOE's Basic Energy Sciences Program.[77] (For a summary, see table 8-7.)

Information is, of course, only one component of decisionmaking, and others may be of far more importance. Joint, cooperative evaluation of projects by researchers and decisionmakers-with participants inside and outside the research area being evaluated--could clarify agency portfolios and researcher needs.[78] Nevertheless, the impacts of some internal agency research may only become known years later.

Research evaluations can help raise difficult questions and uncertainties, but they cannot certify worth. There is simply no convincing way to judge the value of different kinds of research. However, until new techniques, which capture the research process as well as its products, are routinely used, research evaluations cart best be employed to alert agencies to potential successes and problems, and to keep their programs vigilant in research decisionmaking (see table 8-8). Finally, these measurement techniques should be viewed only as one input to agency decisionmaking, because nothing can replace the experienced judgment of program managers and the scientific community to craft a successful research program.

Conclusions

There is a wealth of data on the Federal research system. However, data are most concentrated on Federal R&D funding in universities, degrees awarded in science, the science and engineering work force (especially the Ph.D. component), and some expenditure data by performers. Furthermore, the most detailed analyses are done almost exclusively at NSF and NIH, and not at the other major research agencies. The *highest priority in* data collection for research policymaking in the 1990s is comparable data from all of the agencies, to help Congress maintain a well-rounded view of federally supported research (for a summary, see table 8-9).

This chapter has outlined specific areas in which useful data could be compiled. Specific examples of data on research expenditures, personnel, the research process, and the outcomes of research were detailed. The *second priority are* data presented in forms that are instructive at disaggregated levels of

Table 8-7—Assessment of the Department of Energy's (DOE) Basic Energy Sciences Program: 1982

Objective	Assess the quality of research and performers Estimate the impact of the research on DOE mission Determine program balance Test appropriateness of DOE support
Evaluation questions	Specific scientific problem Research design Findings (past, current, expected) Impact on DOE missions
Methods and data	Ex post peer review Site visits Publication and citation counts Matching peer review (160 reviewers) and bibliometric data Stratified random sample of 125 projects (10 percent of total portfolio worth $250 million)
Recognized constraints	Reviewer variability (no random assignment of reviewers) Sample size too small
outcome	60 percent of projects high overall quality 10 percent of projects exceptional 10 percent low quality
Costs and duration	$700,000-$800,000 Months to complete

SOURCE: U.S. Department of Energy, Office of Energy Research, Office of Program Analysis, *An Assessment of the Basic Energy Sciences Program,* DOE/ER-0123 (Washington, DC: 1982).

76Some projects that were originally funded could have been rejected in hindsight. Likewise, some projects that were originally rejected probably could have delivered reasonable quality. The only way to estimate the quality of projects an agenc rejects for support is to trace its history (which requires the cooperation of agencies and investigators). By examining samples of rejected projects funded by others, some notion of the imperfections of a selection process that led to unwarranted rejection may be obtained. Similarly, how does an agency, or the relevant program within it, determine that a funded project did not meet its stated objectives?

77For other examples, see National Academy of Sciences, op. cit., footnote 8, app. C.

78See J. Jeffrey Franklin, ''Selectivity in Funding: Evaluation of Research in Australia, " *Prometheus,* vol. 6, June 1988, pp. 34-60. The evaluation of an agency *program* would require far more than information on the projects it supports. Rather, questions of implementation-effectiveness and efficiency of decisions, and of the program personnel who make them-would dominate. I short, project evaluations aggregated to the program level would estimate the caliber of researcher performance more than success in administering the program. See Eleanor Chelimsky, ''Expanding GAO's Capabilities in Program Evaluation,'' *The GAO Journal,* winter/spring 1990, pp. 43-52.

Table 8-8—Dimensions of Agency Research Evaluations

Purpose/research question	For example, to fund or not to fund, comparative project performance, extent of contribution to program missions/goals.
Definitions/criteria	For example, quality, priority, cost-effectiveness, innovativeness, success, accountability, impact, productivity, knowledge, growth.
Units of analysis	For example, institute, division, branch/ center, program, project, individual, team, publications, citations, awards, rates of change, adoption/diffusion.
Outcomes	Process v. product, form of research v. content and outputs, cost per-outcome unit, qualitative v. quantitative.
Time horizon	Duration of award, short-v. long-term contribution, continuity/culmination v. new direction.
User audience	Well defined v. fuzzy, disciplinary. multidisciplinary, knowledge- v. problem-oriented.

SOURCE: D.E. Chubin, "Designing Research Program Evaluations: A Science Studies Approach," *Science and Public Policy, vol. 14,* No. *2,* April *1987,* p. *85.*

decisionmaking. In particular, data could be presented to make suitable comparisons and to gauge relative trends (i.e., as indicators of science and technology activity) .79 New indicators, grounded in the tradition of the SEI volumes and extramural research on research, are needed to monitor changes in the Federal research system.[80]

Finally, evaluation techniques of research investments in specific programs and projects were revisited. OTA finds that research evaluation techniques cannot replace mature judgment by policymakers. However, specific evaluation tools, such as bibliometrics and project portfolio analysis, could be further explored. A third *priority is* ongoing project evaluation, which could keep agencies alert to changes in research performance, augment program manager judgments about performers and projects, and serve to improve overall program effectiveness.

In summary, one of the functions of analysis is to raise questions about the information that decisionmakers are currently using to assess their advantages and disadvantages, and to define a richer menu of options. Much information could be collected on the Federal research system to map trends at different levels of aggregation and units of analysis for different users.[81]

However, the existence of data does not ensure its utility, for many policy issues cannot be addressed by additional descriptive information. In particular, external criteria involving the utility of research or impact on objectives can be more persuasive and salient to specific policy decisions. Depending on one's perspective and scope of responsibility, data on budgets, agencies, initiatives, performers, and outcomes can nevertheless clarify understanding of the evolving research system.[82]

This information, however, is not cost-free; nor is the organization for retrieving and distilling it. Congress could consider expanding agency resources to streamline collection and analysis of baseline data. NSF, working in concert with OSTP and OMB, could coordinate and reinforce this national data function, and organizations outside of

79For example, the Engineering Manpower Commission that assembles, analyzes, and disseminates engineering enrollment and degree figures for the American Association of Engineering Societies recently remarked: "There is an old gag among survey researchers that when faced with a choice between consistency and the truth, one should always opt for consistency. When the product of the research is time series data that readers track for years, the virtues of consistency become especially obvious. " See "Consistency Versus Relevance: EMC Changes a Statistic," Engineering *ManpowerBulletin,* June 1990, p. 1. In other words, supplementing a time series with new measures without destroying the continuity of the series is also a virtue.

80Quantitative data will not suffice. Information on the contexts in which research is performed, and characteristics of the performers individually and collectively, will provide clues as to how the numbers can be interpreted and perhaps acted on, For example, see Daniel T.Layzell, "Most Research on Higher Education Is Stale, Irrelevant, and of Little Use to Policymakers, ' *The Chronicle of Higher Education*, vol. 37, No. 8, Oct. 24, 1990, pp. Bl, B3.

81Three decades after historian Derek de Solla Price called for a full-blown "science Of SCienCe, "the policy potential of "research on research" as illustrated in this chapter, has only begun to be exploited. Price's vision is introduced in *Science Since Babylon (New* Haven, CT: Yale University Press, 1961) and *Little Science, Big Science,* op. cit., footnote 21, and elaborated in a series of analyses terminated by his death in 1984. For a retrospective, see Susan E. Cozzens, "DerekPrice and the Paradigm of Science Policy,' *Science, Technology, & Human Values, vol.* 13, Nos. 3 and 4, summer-autumn 1988, pp. 361-372.

82This leads OTA to suggest that the research agencies, especially the National Science Foundation and its policy programs, remain in close touch with analysts of the Federal research system. Keeping abreast of new measurement techniques and findings related to people, funding, and research activities-perhaps through extramural support-wouldbe a modest but fruitful investment in extending inhouse capabilities and refining knowledge of federally sponsored research performance.

Table 8-9-Desired Data and Indicators on the Federal Research System

Category	Description	Method	Primary users: Congress	Agencies	OMB	OSTP
Agency funding allocation method	Funding within and across fields and agencies Cross-agency information on proposal submissions and awards, research costs, and the size and distribution of the research work force supported	Agency data collection (and FCCSET)	X		X	X
Research expenditures	Research expenditures in academia, and Federal and industrial laboratories, centers, and university-industry collaborations Agency allocations of costs within research project budgets, by field Megaproject expenditures: their components, evolution over time, and construction and operating costs	Agency data collection	X	x	x	
Research work force	Size and how much is federally funded Size and composition of research groups	Lead agency survey	X	x		x
Research process	Time commitments of researchers Patterns of communication among researchers Equipment needs across fields (including the fate of old equipment) Requirements for new hires in research positions	Lead agency survey; onsite studies		x		
Outcome measures	Citation impacts for institutions and sets of institutions International collaborations in research areas Research-technology interface, e.g., university/ industry collaboration New production functions and quantitative project selection measures Comparison between earmarked and merit-reviewed project outcomes Evaluation of research projects/programs	Bibliometrics; surveys of industry and academia	x	x		x
Indicators	Proposal success rate, PI success rate, proposal pressure rates, flexibility and continuity of support rates, project award and duration rate, active research community and production unit indices	Agency analysis	X	x		x

KEY: OMB=Office of Management and Budget; OSTP=Office of Science and Technology Policy; FCCSET=Federal Coordinating Council for Science, Engineering, and Technology; PI=principal investigator.

SOURCE: Office of Technology Assessment, 1991.

the government, e.g., NRC and AAAS, could also play critical roles. Refining the measurement process could help to quantify existing opportunities and problems, and pinpoint previously uncovered ones, greatly enhancing research decisionmaking at all levels of the Federal Government.