Research Evaluation, volume 14, number 3, December 2005, pages 000–000, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Disaggregated and flexible

Taxonomy for science and engineering indicators: a reassessment

Mary Kathleen Feeney and Barry Bozeman

Science policy researchers and scientists themselves know reflexively that differences among scientific fields matter. However, sets of government-sponsored science and engineering (S&E) indicators are quite general and in most instances do not report differences among fields. We evaluate the current limitations of S&E indicators, identifying particular data needs about scientific fields. We suggest developing a disaggregated, flexible S&E classification. We argue that disaggregating S&E indicators through faceted analysis or webbed or networked databases will enable users to account for specialization, interdisciplinarity and emerging fields, and to adapt the S&E indicators across databases, making indicators more useful for international comparison.

Mary Kathleen Feeney is Senior Research Associate and Barry Bozeman is Regents' Professor at the Research Value Mapping Program, School of Public Policy, Georgia Institute of Technology, 685 Cherry Street, NW, Atlanta, GA 30330-0345, USA; email: mary.feeney@pubpolicy.gatech.edu; website: www.rvm.gatech.edu

A revised version of this paper was presented at the SRI, Inc. Panel Meeting to Provide Guidance on SRS's Evaluation of its Fields of Science and Engineering Taxonomies in Washington, DC, on 21–22 October 2004. We are grateful to the National Science Foundation, under contract SBR 98-18229, for support of the research project gathering the data reported here.

Wisdom begins by calling things by their right names.

Chinese proverb (anon.)

Our success will depend not on luck, but on forethought.

Wo Fat, *Hawaii Five-0,* broadcast 10 September 1972

FEW PEOPLE, even science and technology policy-makers and analysts, know how many research and development (R&D) laboratories there are in the United States, much less the structural configuration of the laboratories, their environments, and the composition of their R&D (Crow and Bozeman, 1998). Indeed, there is little agreement as to how to define 'R&D laboratory' or 'research center'. One reason for the prodigious gap in our knowledge about the elements of the US National Innovation System is that our models and taxonomies are few in number and, generally, unsophisticated.

Even with much attention to science indicators and productivity measures, taxonomy remains an important science policy enterprise. At the level of institutions, our taxonomic knowledge is limited, but our knowledge of less formal entities, such as science and engineering fields, is in many respects even sketchier. For many reasons, science and engineering fields remain among the most critical elements of our taxonomic understanding. The S&E field is that 'middle range' where the macro-analysis of science policy and the micro-level of analysis of work groups begin to touch one another. If institutional knowledge is important at all, it is important at the level of the scientific field. If the organizational

 0958-2029/05/030000-00 US$08.00

fabric of science policy is in any respect valid, organized as it is around S&E field, then it can be validly informed by analyses of disaggregated field data.

However, there is not so much need to convince scholars, analysts and policy-makers that field is often important, as to understand when it is not important. There is little wisdom in ‘calling things by their right names’ if we do not engage in the sufficient forethought to know that in some instances that right name is ‘engineering’, in others it is ‘chemical engineering’, and in still others it is ‘membranes separation technology’. Adding to complexity, that right name will sometimes not be a field at all but, rather, ‘postdoctoral student’, ‘female’, ‘blue gene computer’ or ‘OMB A21 sponsored research accounting regulations’. To put it another way, it would be useful to develop empirically informed theories about when and why S&E field matters.

Defining scientific disciplines

Sociologists and historians have long debated the most appropriate method for differentiating across scientific knowledge and institutions to define disciplines, specialties and professional units (Cambrosio and Keating, 1983, page 324). According to Whitley (1976) the ‘discipline is an important category for examining scientific elites — and especially their control over training and recruitment’ (page 472). In general, scientific disciplines are defined by unique methods for conducting science and integrating research results (Moore, 1970; Sczacki, 1979), distinct subject matter, knowledge, and insights (Phillips, 1973; Sczacki, 1979), and utility (Phillips, 1973; Blume, 1974). Disciplinary categories divide the scientific labor force, order principles of scientific research and intellectual production, and create a system for allocating prestige and resources (Whitley, 1976; Johnston and Robbins, 1977, page 89).

Whitley contends that scientific disciplines can be subdivided into research areas defined by commitments to differing epistemologies, and further divided into specialties. Unified sets of research practices and techniques define research areas while explanatory models and definitions define specialties. Whitley argues that when explanatory models within specialties become ‘highly developed’ it is critical to define that specialty for ordering scientific ideals (Whitley, 1976, page 474). Chubin (1976) also subdivides disciplines into specialties. However, Chubin defines disciplines based on teaching and university programs and specialties according to research practices.

Historically, specialties and research areas are subunits of larger disciplines implying relationships of power and authority. In some cases, specialties, areas of study and scientific fields expand to claim enough power to become a full discipline. Burr and Leigh (1983) identify seven criteria for discipline development: a unique subject matter; a body of theory and research; a unique methodology; supporting paraphernalia, such as journals and conferences; professional and application utility; a community of scholars; and a belief that the discipline exists. Numerous case studies analyze maturing disciplines and assess the number of papers, personnel, journals and doctoral programs to determine when a specialty becomes a discipline (Griffins and Mullins, 1972; Edge and Mulkay, 1973; Burr and Leigh, 1983). For example, the case for labeling chronobiology as a discipline came with the increase in acceptance of its methods and results and the number of journals, scientific societies and university programs associated with that field (Cambrosio and Keating, 1983). In addition to the above criteria, internal organization, such as collegiate authority and control, and external control, such as industrial and state patronage, of scientific knowledge define the growth, decay and evolution of disciplines (Johnston and Robbins, 1977).

Current classification models categorize science and engineering fields according to hierarchical models, which list entities in relationship to one another, as opposed to networks or webs

Classification models

Whether subdividing science into disciplines, specialties or research areas defined by teaching, research methods or cognitive solidarity, current classification models categorize S&E fields according to hierarchical models as opposed to networks or webs. Hierarchical classification models list entities in relationship to one another, placing the general most inclusive categories (super classes) at the top and the more specific detailed categories (subclasses) at the bottom (Kwasnik, 1999). The super classes are inclusive of all subclasses and require inheritance, where defining factors that are true for a super class hold for the subclasses. For example, S&E fields are often divided into super classes such as life sciences, physical sciences and social sciences. The super class, life sciences, is subdivided into biochemistry, immunology, microbiology and so forth. Hierarchical models are useful because they allow information to flow in two directions, vertically between super and subclasses implying an ‘is-a’ relationship, and horizontally across subclasses.

A variation of the hierarchical model is the tree model, which does not require inheritance, presents classifications that flow in one direction from the super class to the subclass, and represents relationships of power and control, such as chains of command. Tree models are useful because they highlight the relationships and distance between classes and enable users to measure the frequency of entities in the classification (Kwasnik, 1999).

Hierarchical classifications are popular because they present complete and comprehensive information, have an economy of notation and use real definitions. They are ideal for classifying mature domains where the nature of entities and the relationships between those entities are well known. Hierarchical models require that we have full information about the categories in advance of classification and are not appropriate in situations when there are multiple hierarchies or where categories overlap, share attributes or are related. Hierarchical models are particularly difficult to impose when entities do not have the necessary and sufficient criteria to fit into a unique category (Kwasnik, 1999, page 27). For example, some S&E research projects may clearly fit into a specific field or discipline, while others may more or less fit there, but be just as likely to fit somewhere else. Finally, hierarchical models are rigid, forcing new information to fit within the initial design of relationships.

An alternative to hierarchical models is faceted analysis, a flexible classification that is amiable to new information (Ranganathan, 1967). Ranganathan suggests that information be classified according to facets or fundamental categories. For example, we could develop facets for S&E field indicators such as educational training, occupation, knowledge development, techniques and research practices, or type of publication. Individual facets can be internally organized in various types of classifications such as hierarchies, trees or timelines.

Next, the facets are placed in a citation order. For example, we could list educational training as the primary facet and occupation as secondary. Each element in the classification is then assigned a unique citation order that indicates its relationship to other elements in the classification. So two individuals with a PhD in physics would share a primary facet citation (.01) for educational training, but their secondary citation would differ for one who teaches (.01.05) and the other who works in a lab (.01.02).

Facets are a particularly attractive classification model for S&E fields because they can be expanded to include new information. Another benefit is that faceted analysis does not require full information but relies on a common perception of reality. The drawbacks of faceted analysis are that it can be difficult to design, requires knowledge about the needs of potential users, and generally the connections between facets are weak and can be difficult to visualize or represent graphically.

Another alternative method for developing an S&E field taxonomy would be to create webs or networks linking disciplines and subfields. For example, bibliometric analyses measuring co-citation and co-authorship have been used to identify linkages between researchers and centers across disciplines. These studies produce 'maps of science' or graphical representations of these linkages (Small, 1999; Weingart and Stehr, 1999; Morillo, Bordons and Gomez, 2003). Network and faceted analysis, though complex, are ideal for assessing vertical and horizontal linkages.

Importance of taxonomy

Classifying S&E fields enables us to make educational, career and funding comparisons across fields and over time. Of course creating this taxonomy 'inevitably confronts limitations and requires execution of somewhat arbitrary decisions' (National Academies of Sciences, 2004, page 1). As we attempt to develop a more meaningful approach to the taxonomy of S&E fields we must ensure that our indicators are meaningful, comparable, measurable, adaptable to changes and understandable to all stakeholders.

Science and engineering indicators are used for diverse purposes and are important for researchers, students, industry, university administrators, interest groups, and local, national and international policymakers. Indicators enable us to categorize scientific fields (Tijssen, 2001), investigate interdisciplinary relationships (Rinia et al, 2001), and understand collaborative relationships and the social organization of scientists and engineers (Chompalov et al, 2002). Indicators are also used to monitor training and occupational trends in scientific, engineering, technical and mathematical fields in the USA (Fox and Stephan, 2001) and abroad (Beltramo et al, 2001).

Science and engineering indicators are critical to understanding the role of R&D in society. Many studies investigate the links between academic research and industry (Salter and Martin, 2001; Schartinger et al, 2002; Longo et al, 2000; Okubo and Sjoberg, 2000; Jacobbsson, 2002). Van Looy et al (2004) used S&E indicators to understand the relationship between academia and entrepreneurial activity, while McMillan et al (2000) used S&E indicators to study the impact of university research and R&D institutions on the biotechnology industry. Indicators also enable researchers to conduct international comparative R&D studies (Granstrand, 1999; Porter et al, 2002; Fritsch and Franke, 2004)

Science and engineering indicators play a critical role in explicating our understanding of R&D investments and outcomes, and enabling researchers to categorize R&D activities, centers and companies (Forbes and Wield, 2004; Tijssen, 2004; Feller et al, 2002; Feller et al, 2003); understand the changing functions of R&D facilities (Pearce, 1999); and assess the benefits of R&D (Robitaille and Godin, 2002) including knowledge spillovers and innovations from R&D centers (Fritsch and Franke, 2004) and the role of R&D in promoting technological and economic growth (Acs, 1999). Numerous researchers use S&E indicators to analyze patent production (Tijssen, 2001; Liu and White, 2001; Hirschey and Richardson, 2004; Langinier, 2004; Hicks et al, 2001) and the roles of patent production in entry and exit into technical fields (Malerba and Orsenigo, 1999).

Limitations of current indicators

S&E indicators are used to monitor the number of graduates in academic programs, to measure occupational trends in S&E fields and to categorize R&D investments and measure R&D outcomes. Unfortunately, many current users of the National Science Foundation's (NSF) Science Resources Statistics (SRS) data will find that the aggregate S&E fields do not provide them with sufficient information to make comparisons within subfields. For example, researchers are unable to make comparisons within subfields of chemistry or the medical sciences. Students looking to identify a growing sector of the workforce will quickly become frustrated when they are unable to compare the variety of degrees offered at universities with the aggregate S&E data. Despite the importance of S&E field indicators to multiple stakeholders, there are limitations for applications regarding S&E programs, S&E workforce and R&D investments

S&E programs

The first weakness of the current S&E field indicators is that they do not coincide with the number or types of university programs. Programs and degrees offered at universities continue to expand with more specialized fields and interdisciplinary fields that integrate research across various disciplines incorporating knowledge and frameworks from a wide range of organizations and partnerships. For example, Carnegie Mellon University classifies over 10 programs as interdisciplinary, such as computational biology and information technology. At the Georgia Institute of Technology, a PhD in bioinformatics falls under four fields: biology, chemistry and biochemistry, biomedical engineering, and computing. In some cases S&E taxonomies do 'not reflect the organization of graduate programs in many institutions' (National Academies of Science, 2004, page 1). As universities continue to produce graduates in specialized fields it will become more important to have detailed information about academic training and PhD production by disaggregated fields of science and engineering.

As universities continue to produce graduates in specialized fields it will become more important to have detailed information about academic training and PhD production by disaggregated fields of science and engineering

S&E workforce

The classification of the S&E workforce is the second tension within the current taxonomy of field indicators. Today many researchers pursue interdisciplinary, multidisciplinary and cross-disciplinary work, which integrate various disciplinary tools, methods, data, concepts and theories to address complex research questions. The National Science Foundation (NSF) recognizes that identifying individuals by their highest level of education leads to errors when individuals with a bachelor's degree in biology and a PhD in chemistry work primarily with their training in biology. In addition, scientists and engineers working on large projects or at research centers often cross disciplines or use techniques not typically identified with their academic training, making the comparisons of their training and workforce participation weak at best. Finally, the current S&E fields are aggregated at broad levels, which prevents researchers from gathering detailed information about relatively narrow fields of science, such as biophysics, and asking detailed questions about labor trends within broad scientific fields, such as biology.

R&D funding

The third weakness in the current taxonomy of S&E fields emerges when researchers wish to analyze R&D funding. Funding for multidisciplinary and interdisciplinary R&D projects currently falls under the category 'not elsewhere classified' (n.e.c.). The n.e.c. category is continually growing in the NSF Science Resources Statistics (SRS) data. For example, the amount of federal obligations for 'life sciences, n.e.c.' more than quadrupled from 1998 to 2001. The amount of federal obligations for total research to 'engineering, n.e.c.' in 2001 made up more than 25% of the total obligations to all fields of engineering for that year. In the case of psychological sciences, the obligations to 'n.e.c.' were approximately US$668.5 million, more than six times higher than the second grouping 'social aspects' (US$59.8 million) and more than 49 times that of 'biological aspects' (US$13.6 million) (National Science Foundation, 2004b). When catch-all categories become larger than the disaggregate S&E fields, the data become less meaningful to users.

Because SRS data are used for diverse purposes, the data should be disaggregated to a sufficient degree to ensure that multiple users can manipulate the data to fit their specific needs. In order to construct policy-relevant indicators we must identify characteristics of S&E field indicators that are desirable to most, if not all, users. Previous work by Bozeman and Dietz (2001) on constructing policy-relevant indicators for strategic research partnerships suggests an approach thet requires considering the analytical problems of S&E field indicators and the context's user needs. Once we have a conceptual model, which defines the target population and the boundaries of the analysis, we can begin to identify appropriate indicator characteristics. Disaggregated S&E fields should, like all indicators, be policy-relevant, simple, valid, stable, reliable and adequate (Bozeman and Dietz, 2001).

Conceptual model for indicators

The next step is to create a conceptual model for S&E indicators. Though conceptual barriers arise when we attempt to define S&E fields, it is a task for which other US agencies and organizations, such as the National Research Council (NRC), the National Center for Education Statistics (NCES) and the Bureau of Labor Statistics (BLS), have already created taxonomy standards. Disaggregated S&E indicators must: first, be flexible enough to account for interdisciplinary and emerging academic fields; second, capture occupational information; and third, enable researchers to assess R&D funding. Ideally, these indicators will coincide, as well as possible, with S&E field definitions used by other government agencies, universities and international organizations. Finally, our experiences at the Research Value Mapping (RVM) Program lead us to believe that curriculum vitae (CV) analysis may offer a complimentary tool for collecting S&E field data.

S&E programs

S&E field indicators should match as well as possible the types of degrees given at universities. In 2004, the NRC reviewed and approved its current taxonomy for classifying S&E fields. The classification standards define a field by the number of PhD programs, the number of PhDs produced in that field nationally, and the average number of PhDs produced per program, making exceptions for fields included in previous surveys or those that may not fulfill one of the three qualifications because of special circumstances (National Academies of Sciences, 2004, page 19). The NRC taxonomy is a hierarchical model with four super classes: life sciences; physical sciences; mathematics and engineering; social and behavioral sciences; and arts and humanities. The life sciences field is broken into 17 subfields, which are further categorized into 25 super-subfields. The physical sciences, mathematics and engineering field includes 19 subfields and 56 super-subfields. Social and behavioral sciences is broken into 10 subfields and 39 super-subfields, and arts and humanities includes 15 subfields and 54 super-subfields.

Based on user needs, the NSF could expand the NRC classification to include S&E fields defined by undergraduate and masters programs. However, the NSF should use the NRC taxonomy of educational programs with caution when defining S&E careers because it may be the case that individuals work in an interdisciplinary or emerging field, such as information technology, but have training in traditional fields, such as computer science, math, science or business.1

Due to the changing nature of academic disciplines, S&E field indicators must be adaptable over time so that emerging fields have a place in the datasets. An important element of flexibility is anticipating emerging fields. A valuable model for S&E field indicators is the NRC's inclusion of the category 'emerging fields'. S&E field indicators should also include an 'emerging field' category. The inclusion of emerging trends will be invaluable to users who wish to identify growing trends in education and ensure that SRS data remain useful to researchers who are building datasets which incorporate categories for emerging fields.

However, classification models must use categories such as 'emerging fields' with caution because labeling categories as 'miscellaneous' or 'other' is a sign of a premature classification where content is unknown, or relationships are not clearly defined. Kwasnik warns that forcing emerging categories into hierarchical models without complete information can result in 'a representation that is misleading or skewed' (1999, page 28). Classifying emerging S&E research areas, specialties and disciplines may necessitate a model that is more flexible than traditional hierarchical models.

S&E workforce

Given our interest at RVM in analyzing scientific careers and productivity we begin by assuming that useful S&E field indicators must measure the educational training of scientists and engineers, including an appropriate measure for graduates in specialized and interdisciplinary fields. The current S&E indicators define disciplines in hierarchical models based on research strategies, practices and techniques, which prevents researchers from understanding transdisciplinary and interdisciplinary research that may adopt and integrate frameworks and techniques across disciplines. It may be time for SRS to allow S&E workers to check multiple boxes, just as the Census Bureau allows individuals to identify with multiple races and ethnicities. Developing a process by which responses can be weighted across S&E fields may solve problems that arise with interdisciplinary and cross-disciplinary research. Alternatively, SRS could categorize S&E fields using faceted analysis or science maps, which would enable indicators to evolve along with S&E labor markets.

R&D funding

The assessment of R&D investments relies heavily on the use of S&E field indicators. Categorizing S&E fields enables policy-makers to allocate funding by field, researchers to submit proposals based on disciplinary projects, and researchers to monitor and evaluate the outcomes of R&D activities. The usability of SRS data and S&E field indicators for measuring and tracking R&D funding rests primarily on the ability of users to compare these data with other datasets.

'Classifications are never created in a political or social vacuum' (Kwasnik, page 45). Therefore S&E field indicators should align with major national and international datasets to enhance their relevance and usefulness.2 There are three major sources for datasets that classify S&E academic disciplines and occupations in the US: the NRC, the NCES and the BLS. These datasets are similar in that they use hierarchical models to organize S&E field information, but differ in how they define disciplines and subcategories. The NCES and the BLS, which are much larger than the SRS datasets and are defined by the specific needs of their respective users, disaggregate the categories to include fields such as biochemistry, biophysics, computer engineering, biomedical engineering and science technologies (National Center for Education Statistics, 1991). Table 1 indicates that, even at the highest level of classification, these datasets do not currently align.

International

Using discipline and field data from multiple sources requires researchers to crosswalk classifications and sort data into like groups. Combining datasets requires researchers to carefully identify the boundaries and overlaps between categories. These crosswalks inevitably result in the creation of 'not

Using discipline and field data from multiple sources requires researchers to crosswalk classifications and sort data into like groups. Combining datasets requires researchers to carefully identify the boundaries and overlaps between categories

Table 1. Major field categories

Major fields	NRC 2004	OECD 2004	NSF 2004	NSF 1960	NSF 1978	NSF 1953	ISI 2002
Agricultural sciences		X					X
Arts							
Humanities		X					
Arts and humanities	X						
Biology and biochemistry							X
Chemistry							X
Computer science			X**				X
Ecology / environment science			X		X		X
Economics and business							X
Engineering	X*	X	X		X		X
Geosciences							X
Immunology							X
Life sciences	X		X	X	X	X	
Material sciences							X
Mathematics	X*		X**		X		X
Medicine, clinical							X
Medical sciences		X					
Microbiology							X
Molecular biology and genetics							X
Multidisciplinary							X
Natural sciences		X					
Neuroscience and behavior							X
Pharmacology and toxicology							X
Physical sciences	X*		X	X	X	X	
Physics							X
Plant and animal science							X
Psychology / psychiatry			X	X	X		X
Social sciences, general	X	X	X	X	X	X	X
Space science							X
Other sciences, n.e.c.			X	X	X		

Notes: X*: Physical sciences, mathematics and engineering are one category
X**: Mathematics and computer science are one category
NRC: National Research Council, USA
NSF: National Science Foundation, USA
OECD: Organisation for Economic Co-operation and Development
ISI: Thomson Institute for Scientific Information

classified' or 'other' categories, and more importantly may result in miscategorizing or misrepresenting data. For example, sometimes university departments do not align with journals in scientific disciplines, fields and subfields (Bourke and Butler, 1998). For example, looking at Table 2, we see the many challenges one would face aligning two US-based classifications, the NRC and the NSF, under the major fields of 'life sciences'.

However, aligning international datasets does not pose more significant challenges than those one would face crosswalking national datasets. For

Table 2. NRC and NSF taxonomy for life sciences 2004

NRC 2004

Life sciences

- Biochemistry, biophysics, and structural biology
  - Biochemistry
  - Biophysics
  - Structural biology
- Cell biology
- Developmental biology
- Ecology and evolutionary biology
  - Behavior ethnology
  - Biogeochemistry
  - Evolution
  - Population biology
  - Physiological ecology
  - Terrestrial and aquatic ecology
- Genetics, genomics, and bioinformatics
  - Bioinformatics
  - Genetics
  - Genomics
- Immunology and infectious disease
  - Immunity
  - Immunology and infectious disease
  - Immunopathology
  - Immunoprophylaxis and therapy
  - Parasitology
- Microbiology
  - Environmental microbiology and ecology
  - Microbial physiology
  - Pathogenic microbiology
  - Virology
- Molecular biology
- Neuroscience and neurobiology
  - Cognitive neuroscience and neurobiology
  - Computational neuroscience
  - Molecular and cellular neuroscience
  - Systems neuroscience
- Pharmacology, toxicology, and environmental health
  - Environmental health
  - Pharmacology
  - Toxicology
  - Medicinal/pharmaceutical chemistry
- Physiology
- Animal sciences
  - Aquaculture and fisheries
  - Domestic animal sciences
  - Wildlife science
- Entomology
- Food science and engineering
  - Food engineering and processing
  - Food microbiology
  - Food chemistry
  - Food biotechnology
- Nutrition
  - Animal
  - Human, community, and international
- Plant sciences
  - Agronomy and crop sciences
  - Forestry and forest services
  - Horticulture
  - Plant pathology
  - Plant breeding and genetics
- Emerging fields
  - Biotechnology
  - Systems biology

NSF 2004

Life/related sciences

- Agricultural/food sciences
  - Animal sciences
  - Food sciences/technology
  - Plant sciences
  - Other
- Biological sciences
  - Biochemistry/biophysics
  - Biology
  - Botany
  - Cell/molecular biology
  - Ecology
  - Genetics, plant and animal
  - Microbiology
  - Nutritional science
  - Pharmacology, human and animal
  - Physiology, human and animal
  - Zoology
  - Other
- Environmental life sciences
  - Environmental science studies
  - Forestry sciences
- Health/related sciences*
  - Audiology/speech pathology
  - Health services administration
  - Health/medical assistants
  - Health/medical technologies
  - Medical preparatory programs
  - Medicine
  - Nursing, 4 years or longer
  - Pharmacy
  - Physical therapy/other rehabilitation
  - Public health, including environment
  - Other

*(Included under life sciences for doctoral programs only)

example, Bourke and Butler (1998) manually re-categorized the scientific classifications between Thomson ISI, a web-based classification of data on scientists, institutions, countries, and science and technology journals including publications, citations and cites-per-paper counts, and the Australian Standard Research Classification (ASRC). They found that some categories spanned multiple subfields and fields. Their solution to these boundary issues was to retain the ASRC classification, add interfield and interdisciplinary fields, and create a higher category for 'discipline'. In some cases they opted to assign journals to more than one field or subject category. Unfortunately, this careful attempt to crosswalk datasets may have resulted in the misrepresentation or loss of important information.

Aligning the SRS classifications with major international S&E databases, such as the Thomson Scientific ISI, which is commonly used to track patent production, research and development, and workflow across disciplines, and the Organisation for Economic Co-operation and Development (OECD) databases, would enhance the usefulness of the SRS data. The OECD Frascati Manual developed in 1963 and now in its sixth edition, represents the standard for R&D surveys worldwide. The Frascati Manual is used not only by the OECD member countries but also byUNESCO and the European Union, making it an important goal to align this internationally popular dataset with US S&E indicators.

The OECD classifies R&D investments in higher education into six major fields of science and technology: natural sciences, engineering and technology, medical sciences, agricultural sciences, social sciences, and humanities. Those fields are subdivided into 20 categories, of which three are 'other' categories. The OECD classification does not account for interdisciplinary research or enable users to compare R&D data within narrow S&E field categories because the data are not primarily interested in S&E fields, but instead the distribution of R&D by sector (private non-government, government and higher education) and type (basic research, applied research and experimental development). Developing a coding, citation or networked method for classifying S&E fields would make SRS data more flexible for users and compatible with the Thomson Scientific ISI and OECD data, which meet the needs of different types of users.

Research value mapping

Finally, our experiences indicate that disaggregated S&E fields are more useful when investigating field characteristics of large research projects and research centers. Because our work focuses on research centers, which tend to be interdisciplinary and have problem-driven missions, many of the respondents to our surveys work in interdisciplinary fields. In order to capture the evolution of S&E careers over time and across disciplines we have created curricula vitae (CV) databases for scientists and engineers. Our work using CV analysis enables us to capture various levels of training in respondents, giving us numerous opportunities for in-depth analysis of academic and career trends. Our experiences indicate that CV analysis offers a particularly robust approach to gathering data about S&E training and careers.

However, the collection and coding processes are labor-intensive, time-consuming and expensive. CV analysis may provide an alternative method for forward-looking activities, such as defining borders between S&E fields or identifying emerging fields. Limited CV analyses could accompany the data on S&E fields or be used to confirm data. Ultimately, users will find data that further disaggregate S&E fields more useful than the current S&E field taxonomy.

Learned from field-level analysis

To this point, the discussion of disaggregation of fields has itself been at a relatively high level of conceptual 'aggregation'. Perhaps a useful final point is identification of some of the types of information we develop by knowing more about S&E fields. Thus, let us consider some empirical results from the Research Value Mapping Program's recent studies of CV and survey questionnaires data. These data and attendant findings are described fully in a number of RVM publications (e.g. Bozeman and Corley, 2004; Lee and Bozeman, in press; Dietz, 2004; Dietz and Bozeman, in press; Lin and Bozeman, in press; Gaughan and Bozeman, 2002). Our purpose here is not to reiterate those findings but, instead, to simply show some of the empirical connections derived from analysis of field-level data. Just to give a flavor for the types of dynamics one can identify from analysis of fields, here are some miscellaneous but perhaps suggestive results based on a simple correlational analysis (all results reported below are based on descriptive data or on correlations at 0.01 or greater level of statistical significance):

- Compared to all other fields, *biologists* are least likely to report and work with industry.
- *Computer scientists*, on average, support the largest number of graduate students, biologists the least.
- When examining motivations for research collaboration, *electrical engi*neers are most likely to be motivated by a desire to mentor students and postdoctoral researchers, *chemists* are the least likely to be so motivated.
- *Chemists* and *electrical engineers* have the largest number of total collaborators.
- *Biologists* have the highest percentages of female collaborators and *chemists* and *electrical engineers* the least.

- *Civil engineers* are most likely to report a desire to transition into an industry job.
- *Chemists* tend to have the most patents.
- There are no significant differences among fields with respect to:
  1. reported level of job satisfaction;
  2. importance or concern about job security;
  3. marital rates or children;
  4. 'cosmopolitanism' of research collaboration (e.g. people outside one's university).

Interestingly, it is the policy-relevant variables that show the greatest difference among fields and the personal variables (e.g. job satisfaction, political and social attitudes, views about one's institution) that show the least. This implies, perhaps, that field disaggregation is especially important for those indicators of greatest interest to policy-makers. However, the findings above are just appetizers, suggesting the potential value of data about S&E field. In all likelihood, taxonomic advance in our understanding of S&E fields will generate not only increased understanding of fragmented empirical findings but, more importantly, will accelerate our theoretical knowledge of the implications of fields for behavior and outcomes.

Notes

1. The majority of information technology (IT) jobs are held by people with degrees in math, science, social science and business, with only 29% of IT workers holding degrees in computer science (US General Accounting Office, 1998, cited in National Academies of Sciences, 2000a).
2. National Academies of Science (2000b) argues that the NSF 'should coordinate definition and categories across agencies to facilitate a consistent picture of the different stages in the market, from student training and degree choice to mid-career transitions across and out of science and engineering fields' (page 5).

References

Z J Acs (1999), 'Public policies to support new technology-based firms (NTBFs)', *Science and Public Policy*, pages 247–257.

J P Beltramo, JJ Paul and C Perret (2001), 'The recruitment of researchers and the organization of scientific activity in industry', *International Journal of Technology Management*, 22(7–8), pages 811–834.

S S Blume (1974), *Toward a Political Sociology of Science* (Free Press, New York, USA).

P Bourke and L Butler (1998), 'Institutions and the map of science: matching university departments and fields of research', *Research Policy*, 26(6), pages 711–718.

B Bozeman and E Corley (2004), 'Scientists' collaboration strategies: implications for scientific and technical human capital', *Research Policy*, 33(4), pages 599–616.

B Bozeman and J Dietz (2001), 'Strategic research partnerships: constructing policy-relevant indicators', *Journal of Technology Transfer*, 4, 26 October, pages 385–393.

W R Burr and G K Leigh (1983), 'Famology: a new discipline', *Journal of Marriage and the Family*, 45(3), August, pages 467–480.

A Cambrosio and P Keating (1983), 'The disciplinary stake: the case of chronobiology', *Social Studies of Science*, 13(3), August, pages 323–353.

I Chompalov, J Genuth and W Shrum (2002), 'The organization of scientific collaborations', *Research Policy*, 31(5), July, pages 749–767.

D E Chubin (1976), 'The conceptualization of scientific specialties', *Sociological Quarterly*, 17, pages 448–476.

E Corley, B Bozeman and M Gaughan (2003), 'Evaluating the impacts on grants on women scientists' careers: the curriculum vita as a tool for research assessment', in P Shapira and S Kuhlmann (editors), *Learning from Science and Technology Policy Evaluation: Experiences from the US and Europe* (Edward Elgar, Cheltenham, UK).

M Crow and B Bozeman (1998), *Limited by Design: R&D Laboratories in the US National Innovation System* (Columbia University Press, New York, USA).

J Dietz (2004), 'Scientists and engineers in academic research centers: an examination of career patterns and productivity', unpublished PhD dissertation, School of Public Policy, Georgia Institute of Technology, Atlanta, Georgia, USA.

J Dietz and B Bozeman (in press), 'Academic careers, patents, and productivity: industry experience as scientific and technical human capital,' *Research Policy*.

I I Feller, C P Ailes and J D Roessner (2002), 'Impacts of research universities on technological innovation in industry: evidence from engineering research centers', *Research P*olicy, 31(3), March, pages 457–474.

I Feller, G Gamota and W Valdez (2003), 'Developing science indicators for basic science offices with mission agencies', *Research Evaluation*, 12(1), April, pages 71–79.

N Forbes and D Wield (2004), 'What is R&D? Why does it matter?', *Science and Public Policy*, 31(4), August, pages 267–278.

M F Fox and P Stephan (2001), 'Careers of young scientists: preferences, prospects and realities by gender and field', *Social Studies of Science*, 31(1), February, pages 109–122.

M Fritsch and G Franke (2004), 'Innovation, regional knowledge spillovers and R&D cooperation', *Research Policy*, 33(2), March, pages 245–255.

M Gaughan and B Bozeman (2002), 'Using curriculum vitae to compare some impacts of NSF research center grant with research center funding', *Research Evaluation*, 11(1), pages 17–26.

B Godin (2004), 'The obsession for competitiveness and its impact on statistics: the construction of high-technology indicators', *Research Policy*, 33(8), October, pages 1217–1229.

O Granstrand (1999), 'Internationalization of corporate R&D: a study of Japanese and Swedish corporations', *Research Policy*, 28(2–3), March, pages 275–30.

B C Griffins and N C Mullins (1972). 'Coherent social groups in scientific change', *Science*, New Series, 177(4053), 15 September, pages 959–964.

D Hicks, T Breitzman, D Olivastro and K Hamilton (2001), 'The changing composition of innovative activity in the US: a portrait based on patent analysis', *Research Policy*, 30(4), April, pages 681–703.

M Hirschey and V J Richardson (2004), 'Are scientific indicators of patent quality useful to investors?', *Journal of Empirical Finance*, 11(1), pages 91–107.

S Jacobbsson (2002), 'Universities and industrial transformation: an interpretative and selective literature study with special emphasis on Sweden', *Science and Public Policy*, pages 345–365.

R Johnston and D Robbins (1977), 'The development of specialties in industrialized science', *Sociological Review*, 25, pages 87–108.

B H Kwasnik (1999), 'The role of classification in knowledge representation and discovery', *Library Trends*, 48(1), Summer, pages 22–47.

C Langinier (2004), 'Are patents strategic barriers to entry?', *Journal of Economics and Business*, 56(5), Sept–Oct, pages 349–361.

S Lee and B Bozeman (in press) 'The effects of scientific collaboration on productivity', *Social Studies of Science*.

M-W Lin and B Bozeman (in press) 'Researchers' industry experience and productivity in university-industry research centers: a scientific and technical human capital explanation', *Journal of Technology Transfer*.

X L Liu and S White (2001), 'An exploration into regional variation in innovative activity in China', *International Journal of Technology Management*, 21(1–2), pages 114–129.

P E Longo et al (2000), 'Reengineering engineering research and education in Brazil: cooperative networks and coalitions', *Science and Public Policy*, pages 37–44.

F Malerba and L Orsenigo (1999), 'Technological entry, exit and survival: an empirical analysis of patent data', *Research Policy*, 28(6), August, pages 643–660.

S G McMillan, F Narin and D L Deeds (2000), 'An analysis of the critical role of public science in innovation: the case of biotechnology', *Research Policy*, 29(1), January, pages 1–8.

W E Moore (1970), *The Professions: Roles and Rules* (Russell Sage, New York, USA).

F Morillo, M Bordons and I Gomez (2003), 'Interdisciplinarity in science: a tentative typology of disciplines and research areas', *Journal of the American Society for Information Science and Technology*, 54(13), pages 1237–1249.

M J Mulkay and D O Edge (1973), 'Cognitive, technical ad social factors in growth of radio astronomy', *Social Science Information sur Les Sciences Sociales*, 12(6), pages 25–61.

National Academies of Sciences (2000a), *Building a Workforce for the Information Economy* (National Academies Press, Washington DC, USA).

National Academies of Science (2000b), *Forecasting Demand of Supply of Doctoral Scientists and Engineers: Report of a Workshop on Methodology* (National Academies Press, Washington DC, USA).

National Academies of Sciences (2004), *Assessing Research-Doctorate Programs: A Methodology Study* (National Academies Press, Washington DC, USA).

National Center for Education Statistics, US Department of Education (1991), *Classification of Instructional Programs — CIP, Series, Code and Title Listing — 1990 Update*. Integrated Postsecondary Education Data System, July (Washington DC, USA).

National Science Foundation, Division of Science Resources Statistics (2004a), *Federal Funds for Research and Development: Fiscal Years 2001, 2002, and 2003*, NSF 04-310, Project Officer, Ronald L Meeks (Arlington, VA, USA).

National Science Foundation, Division of Science Resources Statistics (2004b), *Federal Obligations for Research by Agency and Detailed Field of Science and Engineering: Fiscal Years 1970–2002*, NSF 04-313, Project Officer, Ronald L. Meeks (Arlington, VA, USA).

Y Okubo and C Sjöberg (2000), 'The changing pattern of industrial scientific research collaboration in Sweden', *Research Policy*, 29(1), January, pages 81–98.

R D Pearce (1999), 'Decentralised R&D and strategic competitiveness: globalised approaches to generation and use of technology in multinational enterprises (MNEs)', *Research Policy*, 28 (2–3), March, pages 157–178.

D K Phillips (1973), 'Sociologists and their knowledge', in E Freidson (editor), *The Professions and Their Prospects* (Sage Publications, Beverly Hills, CA, USA).

A L Porter et al (2002), 'Measuring national "emerging technology"', *Science and Public Policy*, pages 189–200.

S R Ranganathan (1967), *Prologomena to Library Classification*, 3rd edition (Asia Publishing House, Bombay, India).

E J Rinia, T N. van Leeuwen, HG van Vuren and A F J van Raan (2001), 'Influence of interdisciplinary on peer-review and bibliometric evaluation in physics research', *Research Policy*, 30(3), March, pages 357–361.

J-P Robitaille and B Godin (2002), 'Evaluation of governments' scientific output: a bibliometric profile of Canada', *Science and Public Policy*, pages 59–68.

A K Salter and B R Martin (2001), 'The economic benefit of publicly funded basic research: a critical review', *Research Policy*, 30(3), March, pages 509–532.

D Schartinger, C Rammer, M M Fischer and J Frolich (2002), 'Knowledge interactions between universities and industry in Austria: sectoral patterns and determinants', *Research Policy*, 31(3), March, pages 303–328.

J Sczacki (1979) *History of Sociological Thought* (Greenwood, Westport, CT, USA).

H Small (1999), 'A passage through science: crossing disciplinary boundaries', *Library Trends*, 48, pages 72–108.

R J W Tijssen (2001), 'Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows', *Research Policy*, 30(1), January, pages 35–54.

R J W Tijssen (2004), 'Is the commercialisation of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles', *Research Policy*, 33(5), July, pages 709–733.

US General Accounting Office, Health, Education, and Human Services Division (1998), *Information Technology: Assessment of the Department of Commerce's Report on Workforce Demand and Supply*, GAP/HEHS 98 106R. Washington DC, 20 March 20.

B van Looy et al (2004), 'Combining entrepreneurial and scientific performance in academia: towards a compounded and reciprocal Matthew effect?', *Research Policy*, 33(3), April, pages 425–441.

P Weingart and N Stehr (editors) (1999), *Practising Interdisciplinarity* (University of Toronto Press, Toronto, Canada).

R Whitley (1976), 'Umbrella and polytheistic scientific disciplines and their elites', *Social Studies of Science*, 6(3/4), Special Issue: Aspects of the Sociology of Science: Papers from a Conference, University of York, UK, 16–18 September, pages 471–497.