WORKSHOP ON DEVELOPING A NEW RESEARCH DATA INFRASTRUCTURE FOR THE STUDY OF ORGANIZATIONS AND INNOVATION

NATIONAL SCIENCE FOUNDATION HEADQUARTERS
ARLINGTON, VIRGINIA
JULY 23, 2008

WORKSHOP REPORT

CAROL CORRADO
SENIOR ADVISOR AND RESEARCH DIRECTOR,
THE CONFERENCE BOARD | ECONOMIC PROGRAMS

November 10, 2008

Table of Contents

Introduction

Concern that our nation may be under-investing in innovation and in the factors that give rise to innovation is growing.[1] This concern is not based on empirical evidence because few effective measures of innovation are available, resulting in little capacity to identify and effectively predict how investments in innovation and its infrastructure will affect our nation's competitiveness. As part of a broader strategy to augment this capacity, the National Science Foundation (NSF) sponsored a workshop organized by The Conference Board to identify a research agenda that could be used to build an empirical platform for the scientific study of innovation within organizations.

The potential to build such a platform has opened up for several reasons. First, the business community recognizes they need to measure and promote innovation within their own organizations. Innovation is at the top of most corporate agendas, and interest in developing firm level innovation metrics is intense. Second, fundamental advances in the collection and analysis of new types of data on social activity opens possibilities ranging from scraping the web to capturing click streams to videoing team interactions. Third, advances in confidentiality protection now make possible the analysis of sensitive business data without revealing the identities of the respondents—enabling the scientific study of the innovation process. Finally, policy-makers seek ways to "grow our economy" through innovation, but analysts lack indicators and tools to advance the public discourse in the United States.

The approach was to bring together a small group of business representatives with experts who study organizations, data collection, and confidentiality issues. The workshop was an informal venue whereby the group identified promising research areas and gaps in data needed for the study of organizations and innovation.

The experts at the workshop focused on the goal of *measuring innovation* via developing a broad understanding of innovation's processes, lifecycles, and role in the economy and global business environment. The ultimate goals were to better understand the *drivers and consequences* of innovation.

Taking Stock

The group viewed innovation broadly, namely, as a dynamic process by which organizations put something new to commercial use or financial gain—a view consistent

[1] For example, see the reports of the National Academies' *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (Washington, DC: National Academies Press, 2007), the Council on Competitiveness' *Innovate America* (available at http://www.compete.org, 2005), and the President's Council of Advisors on Science and Technology's *Assessing the U.S. R&D Investment* (http://www.ostp.gov/cs/pcast/documents_reports, 2002).

with the definition used in the report to the Secretary of Commerce by the Advisory Committee on Measuring Innovation in the 21st Century Economy.[2]

The group pointed to major gaps in our current data systems and knowledge base. In particular, the measurement of the *net outcome of innovations* – the impact on economic activity – is inadequate. In addition, little is known about the *characteristics of innovations* in terms of the resultant success or failure. The gaps are evident at all levels of analysis, that is, organizational, regional, national, and multinational.

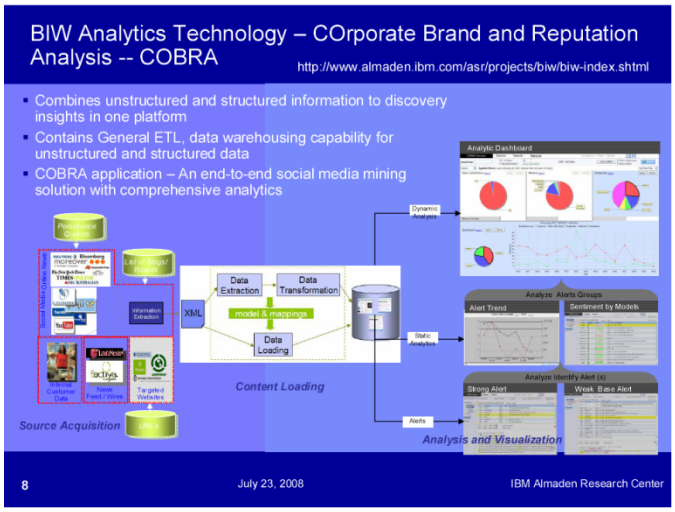

Figure 1: Source—Jeanette Blomberg, IBM Almaden Research Center.

The group identified a number of exciting new advances that could be built on to fill those gaps. For example, work at IBM (Figure 1) shows that it is possible to mine the unstructured databases generated from team interactions for analytical purposes. Other work at Cornell University uses a web archive to study the evolution and growth of business by tracking their web presence (Figure 2).

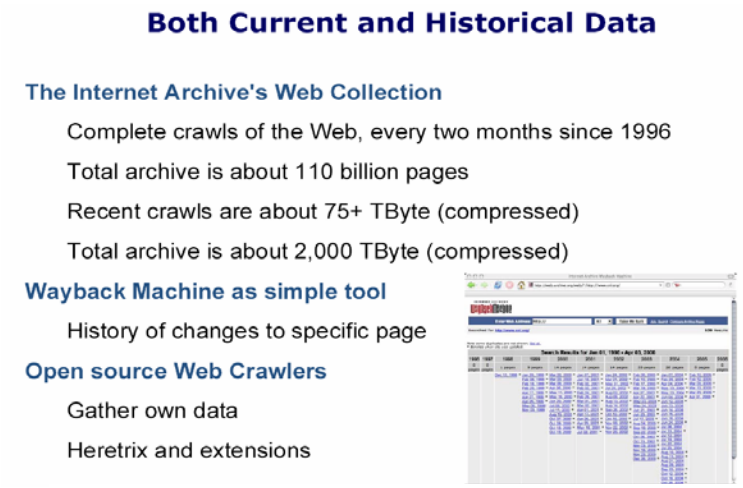

Figure 2: Source—Dan Huttenlocher, Cornell University.

The group recognized that in developing new data for the study of innovation, nothing is more fundamental than considering who and what is to be measured. As previously noted, the experts considered innovation broadly (i.e., beyond S&E) and mainly within a business context, although there was also discussion of the university, community, and international aspects. The group identified four key sets of questions that a dataset on organizations could be used to address.

[2] Available at http://www.innovationmetrics.gov.

The first of these is to develop a scientific way of measuring *what* innovation is: identifying the units, the scales, and the level and trajectory of activity. This is the first step to determining how economic value gets created through innovation. The group thought that determining the rate of return to projects within a company by developing ways of measuring the full cost of the inputs to innovation (over its complete lifecycle) and the business outcomes expected and/or actualized from those inputs was a priority. Such an approach would permit businesses to better understand the distribution of rates of return for different projects, including appropriate time horizons.

The second is to advance an understanding of *how* and *why* innovation takes place. This is necessary to identify the inputs to innovation (including knowledge itself), the determinants of successful innovations, and the factors that affect how innovations diffuse, such as social networks and geography. Characterizing the features and practices of organizations, individuals, industries, markets, and nations—and the links among them—that promote innovation (including the skill/talent/training of the workforce) is necessary.

The third is to understand the *consequences* of innovation. This is particularly true in terms of understanding the impact of innovation on aggregate economic activity, but also the effect of outsourcing of parts of value chain in terms of economic vulnerability of locations, unemployment, the associated political outcomes, and corporate social responsibility.

The fourth is to understand the broader *environment*: globalization, technological change, and innovation are interdependent processes in our economy. New ways of communicating exemplified by Web 2.0 will change business's customers, suppliers and partners. Trends in emerging markets and competitiveness will determine the pattern of global engagement. A better understanding of the impact of formal and informal interactions on the boundaries of companies and industries, the diffusion of technology and ideas, and the larger process of value creation can inform business strategy and policy-making.

Findings

The workshop participants recommended that a new research data infrastructure be created to foster transformative research on innovation. The approach is summarized in four findings.

The first two findings identify *new directions in business data creation/collection and innovation research.* One approach is research grounded in direct observation and suggests that data on the *net outcomes* of innovation projects could be collected from selected companies and used to study the determinants and characteristics of innovation successes and failures. The second is a "micro-to-macro" approach which suggests comprehensive data on *innovation input costs* could be collected according to concepts used in modern business organizations. These two approaches fill the major information gaps as seen by participants at the workshop.

The second two recommendations identify *key steps*. A new research data infrastructure could develop links to existing data and research on other key organizations/entities in the innovation process, such as universities, entrepreneurs, and affiliates in foreign countries. A broad-based effort thus could be mounted to support the building of an empirical platform for innovation research.

Further elaboration of these findings follows:

Finding 1:
The creation of basic data on *innovation net outcomes* would allow the study of the determinants and characteristics of innovation successes and failures. Such data would permit the extension of current work (see, for example, Figure 3) to examine the roles of organizational practices (employment and management), organizational characteristics (employee knowledge and skills, business model, IT use), environmental and cultural factors (location and networks), entrepreneurial factors (firm age and origin), as well as other emerging critical factors and dynamics.

EXTERNAL VALIDATION: BETTER PERFORMANCE IS CORRELATED WITH BETTER MANAGEMENT

Dependent variable	Labor Productivity	Profits (ROCE)	Sales growth	Share Price (Tobin Q)	Exit
Estimation	OLS	OLS	OLS	OLS	Probit
Firm sample	All	All	All	Quoted	All
Management	**0.198*****	**1.880****	**0.032****	**0.250*****	**-0.200****
Firms	2706	1752	2145	374	709

Figure 3: Source—Nick Bloom, Stanford University.

Identify and study *innovation processes within organizations* by collecting data on inputs and outcomes of innovation *projects* within organizations. The project is the basic unit of production in most services firms and an identifiable segment of the value chain in many others. In still others, determining and developing this "unit of innovation" is an important subject of research for both participating companies and organizations more generally.

Finding 2:
Headway has been made in classifying and collecting data according to business functions. These concepts were found to have substantial meaning and applicability for respondents to business surveys (Figure 4). The group found that national totals of business spending on selected functions could potentially form the basis of new indicators of business investment in innovation (e.g., the sum

Figure 4: Source—Clair Brown and Tim Sturgeon, University of California (Berkeley) and MIT.

of spending on "new product and/or service development" and "strategic management/new process development").

Data on selected economic and social outcomes based on a sample of the universe of employer firms thus could be collected according to *business functions*. The resulting longitudinal microdata, if appropriately designed, could produce a vehicle that could be used to standardize and benchmark findings from exploratory project-level innovation process research. The vehicle could also be used to collect public research data for the macroeconomic analysis of business and policy questions of the day.

Finding 3

Existing research finds that the innovation process often is long and complex and increasingly open; that the origin of innovation may be in entrepreneurship, a commercial ally, or work done in university and government-sponsored laboratories; and that the reach of innovation within an organization is not bound by geography.

Data on *innovation processes that span organizations and locations* (alliances, universities, foreign affiliates) could be identified, collected, and studied. Information on the multiple entities (including individuals) that are needed to cover and study complete innovation lifecycles could be generated as part of research on the origin and global dynamics of innovative activity.

Finding 4

The group found that new cyber tools and advances in confidentiality have transformed the way in which data can be collected within businesses. The group also found that existing sources could be leveraged to develop data on innovation. Finally, the group found that businesses saw major incentives to participating in developing new data on innovation: They could learn from the participating scientists and build a broader understanding of innovations across other business and global cultures.

Research projects that focus on developing information from innovative companies and on innovation projects within those companies thus could benefit from the participation of computer and information scientists. For example, the work to develop project-level innovation process data could use and further develop new approaches to data collection (that is, testing noninvasive techniques to harvest data from within firms and auto populating researcher databases). More generally, insofar as opportunities in a broad range of applications are created, the bringing together of scientists to consider business microdata privacy/access and data collection from organizations could itself be significant. In sum, interdisciplinary teams of sponsored social and computer science researchers (SBE and CISE) working with business strategists and research-oriented practitioners to measure and study innovation processes could yield new models, tools, and indicators of innovation.[3]

[3] SBE and CISE refer to the NSF's Social, Behavioral, and Economic Sciences and Computer and Information Science and Engineering Directorates, respectively.

Next Steps

The creation and analysis of representative information are core elements of the scientific endeavor. No less fundamental is the need to replicate analysis and protect respondent identity. For a variety of reasons, currently available business microdata generally do not jointly meet these criteria of scientific inquiry (representative coverage, researcher access/replicability, and confidentiality protection). Developing the needed innovation business microdata and research along with an infrastructure for access and protection requires solving a series of technical and social challenges.

There is broad interest in such a research program. The NSF has programmatic support for the advancement of research on innovation and organizations and is also required to respond to the America COMPETES Act. The NSF is uniquely poised through its CISE directorate to draw upon data collection and confidentiality experts (computer and information scientists, statisticians) to develop the needed safeguards and access tools.

For more than ninety years, The Conference Board has found that leaders in businesses, nonprofit organizations, and NGOs learn and innovate based on information they gather from each other and glean from economic and management experts. The Kauffman Foundation has deep interests in promoting research and policy that help to better understand entrepreneurship. And The Alfred P. Sloan Foundation has deep interests in the functioning of economic institutions, behavior and performance. These communities have a rich history of collecting unique data from a variety of sources and using the data to develop empirically-based understanding of innovation and productivity.

A national research infrastructure for the study of organizations supported by these important groups would engage the business community relative to alternative approaches in many ways. Vital to this engagement—and success of the overall effort—is that business play a crucial role in framing the content of the data collection effort along with the researchers. Equally vital is that the underlying sensitivities of the business community regarding confidentiality and protection be accurately garnered and fully addressed in the data structure and access tools. *This report's recommendations for new directions and requirement of scientific inquiry meet, perhaps uniquely, these needs and sensitivities of the business community.*

The business and scientific communities alike will see the establishment of such a national research data infrastructure for the study of organizations as both evolutionary and transformational. It would represent the coming of age of "management science" and of evidence-based social and economic research at the firm/company-level. Business strategy will be better informed through business participation—and the better data, metrics and research on innovation will be available to address national concerns.

* * * * * * * * * * * * * * *

The remainder of this document presents the detailed report from the workshop, which was an all-day workshop consisting of six sessions. Opening and closing plenary sessions sandwiched four others in which (1) experts in data collection, data access, and confidentiality addressed all participants, and (2) participants were split into three breakout working sessions. In the working sessions, participants were asked to first pose research questions and then to specify the data needed for researching the answers and how that data could be created/collected. The workshop agenda, short biographies of participants, list of all attendees (including invited observers) is attached as an Appendix.

The detailed report reviews the background for the workshop and the results of the workshop. It outlines the potential, as identified by the experts at the workshop, for the NSF to develop new data and a new data infrastructure for research on innovation and organizations. It has been circulated for comment (see the Acknowledgments section) and will be the subject of a luncheon panel discussion the *2008 Kauffman Symposium on Entrepreneurship and Innovation Data*, to be held November 21 in Washington, D.C. (http://www.kauffman.org/dataSymposium/2008index.cfm).

* * * * * * * * * * * * * * * *

Workshop Background

In modern economies, economic value is derived increasingly through the making and selling ideas. Less than three centuries ago, the primary basis of economic value in human society was the production and trade of food—and social, behavioral and economic thought was grounded in the world of agriculture. The Industrial Revolution created a new social scientific infrastructure: Human beings could now add value by making and selling things other than food, and new social and economic theories, new behavioral models, and new data and ways of collecting that data (largely on manufacturing firms and workers) emerged.[4]

The scientific challenge of today is to advance our broad understanding of economic and social value creation through innovation and knowledge appropriation. Building an empirical platform for the scientific study of the nation's engines of innovation, economic value creation, and competitiveness—organizations—is a major step forward in meeting that challenge.

Despite the importance and relevance of improving the empirics of organizations, for a number of reasons, no database exists that is widely available to researchers to examine firm behavior. The consequences of this lack of data are profound: Policy on topics such as offshoring, technological change, and workforce productivity and competitiveness is often informed by anecdote rather than by evidence. Evidence that is available may not be based on representative samples of data. And, although a community of practice encompassing the study of business microdata is in its infancy, graduate students and junior researchers currently are not sufficiently trained in the complexities of business definitions, dynamics, and data creation to fully participate.

New needs and previous approaches

Existing data infrastructures are not sufficient for researchers to model, measure, and study the evolving mechanisms whereby innovating enterprises and entrepreneurs create economic value. The call for better data and metrics on innovation was made clear by the America COMPETES Act, the report to the Secretary of Commerce by the Advisory Committee on Measuring Innovation in the 21st Century Economy, and the National Academies' report on Understanding Business Dynamics (a panel of the Committee on National Statistics).[5]

Business activity is the basic engine of innovation and economic growth, creating jobs and generating income. Although a large empirical literature has yielded insights into topics that fundamentally affect the business environment (such as taxation, regulation,

[4] The ferment is well described by Robert Heilbroner in, *The Worldly Philosophers: the lives, times, and ideas of the great economic thinkers* (New York, N.Y.: Touchstone Press, 7th edition 1995).
[5] John Haltiwanger, Lisa M. Lynch, and Christopher Mackie, eds., *An Integrated Data System for America's Future* (Washington, D.C.: The National Academies Press 2007).

and technical change), the underlying mechanisms that generate entrepreneurship and foster the innovation process within organizations are not well understood. Until innovation and entrepreneurship are better measured, modeled, and studied from both within and outside of business organizations—and a more or less commonly accepted body of scientific knowledge emerges—policy formulation, business attitudes, and academic research will remain disconnected.

Several approaches have been taken to create business datasets that researchers can use to increase the scientific understanding about innovation and organizational change. One approach was a partnership between academics and businesses that developed a business database called the PIMS project (Profit Impact of Marketing Strategy). This project created a large panel dataset of firms and provided new insights into business decisions such as market entry, pricing and product quality. This project fell into disuse for a variety of reasons, however, and little academic research has used the data in recent years. Nonetheless, the PIMS project is an example of applied research that pushed the frontiers of business strategy formulation.[6]

Another approach, partially supported by the National Science Foundation, is to provide access to the Census Bureau's Business Register by permitting researchers to work with the data at eight Research Data Centers. The resulting research has generated new insights into firm behavior, job creation and job destruction. A related infrastructure project was the Longitudinal Employer-Household Dynamics (LEHD) program which provided, for the first time, an infrastructure that could analyze the impact of economic turbulence on worker job ladders, career paths and firm performance. These data are not widely used, however, as access costs several thousand dollars a month, the process of proposal approval is arduous, and researchers must travel to one of the eight Data Center sites.

Other approaches have turned to commercial datasets, such as Standard & Poor's COMPUSTAT and the files made available by the Center for Research on Security Prices (CRSP) or Wharton Research Data Services. The availability of these files, which provide financial and accounting information on publicly traded companies, has had a major influence on financial and accounting research. Similarly, datasets like Dunn and Bradstreet and ABI/Inform are often used as sample frames for nongovernmental surveys. However, getting representative and relevant research data from commercial sources is difficult. COMPUSTAT and CRSP are able to cover publicly-held companies only, and the content is largely aimed at serving institutional investors; the Dunn and Bradstreet and ABI/Inform datasets are primarily for marketing purposes. As a result, there are substantial problems with these datasets, and their use for supporting a broader research agenda is highly questionable.

[6] See papers and analysis in Farris, Paul W. and Michael J. Moore, eds., *The Profit Impact of Marketing Strategy Project: Retrospect and Prospects* (Cambridge, U.K: Cambridge University Press 2004).

The international community has approached the lack of statistical information on innovation (beyond existing R&D and patent indicators) by developing the Community Innovation Survey (CIS). The CIS has been collected widely throughout Europe and other countries, such as Australia and Japan, since the early 1990s. Though the CIS has become institutionalized, evidence based on CIS surveys has yet to significantly influence the development of policy (in any country with a CIS, much less the United States, which does not have a CIS). The available scientific findings to date are apparently somewhat limited, and they and the available CIS metrics have not been oriented to answering questions that are relevant to policy formulation.[7] Moreover, the surveys reportedly have low response rates.

Protecting confidentiality and providing access

The creation and analysis of representative information are core elements of the scientific endeavor. No less fundamental is need to protect respondent identity (confidentiality) and the ability to replicate scientific analysis (access). The development of the needed business microdata, access infrastructure, and high quality research thus requires solving a series of technical and social challenges, namely:

- The organizations that are the sources of the data must have their information protected from access and use by unauthorized individuals and for unauthorized purposes, and must be convinced that this protection is in place.
- The custodians of the data on organizations must have taken, and be confident that they have taken, all reasonable means to protect the data.
- Researchers must be provided with a research environment that facilitates high-quality research.
- The benefits of researcher access to microdata must be clearly demonstrable to justify both the risk and the cost of providing that access.

Potential for change

Without an investment in data on organizations that engages (1) use by the best researchers, and (2) the participation of business, while providing necessary the safeguards and researcher access, the present situation is unlikely to change.

The potential for getting such information through federal statistical agencies is very small: New surveys, even new questions and modules on existing, representative surveys can take up to a decade. Capturing high quality information about key measures, such as technology and personnel practices, is difficult both because of respondent burden and problems with identifying the right respondent. In addition, data collected by federal statistical agencies are often not well suited for amendment because they are collected for "core" statistical purposes — the Census Bureau's business data collection is primarily structured for our national accounts while the Bureau of Labor Statistics' data programs are designed to provide information about labor markets and prices.

[7] Anthony Arundel, "Innovation Survey Indicators: Any progress since 1996?" available at http://www.oecd.org/dataoecd/24/28/37436234.pdf

There is clear potential for NSF/SBE to play a lead role in developing a research data infrastructure that will respond to the innovation information gap. Since the early 1970's, recognizing the importance of studying individual, family and political behavior, NSF has supported the American National Election Studies program, the Panel Study of Income Dynamics, and the General Social Survey (GSS). In addition to the programmatic fit as a response to the national imperative of the America COMPETES Act, NSF/CISE is uniquely poised through its CyberTrust program to play a lead role in creating the needed safeguards while providing access to confidential business data.

For more than ninety years, The Conference Board has found that leaders in businesses, nonprofit organizations, and NGOs learn and innovate based on information they gather from each other and glean from economic and management experts. A national research infrastructure for the study of organizations offers a new and exciting means of bringing the collective wisdom of leaders and the experience of organizations to bear on business, economic and social concerns in a scientific manner.

A national research infrastructure for the study of organizations supported by the NSF would engage the business community and best researchers relative to previous approaches insofar as:

- The above-mentioned challenges of ensuring confidentiality and providing access to the data and information are met.
- The data and information collected is based on representative samples and is at the appropriate level of analysis: the company/firm (or organization).
- The data and information collected are well suited for addressing timely and relevant issues.

These considerations imply that: the business community must play a crucial role in framing the content and structure of data collection effort; the underlying sensitivities of the business community regarding confidentiality and access must be accurately garnered and fully addressed by the structure; and those sensitivities are understood and respected by the researchers who access the data.

Workshop Results

The measurement and understanding of innovation and innovation policy is limited by major data gaps. The workshop agenda was designed to challenge the participants to frame the content, structure, and approach to building an infrastructure that could fill those gaps. The opening session included speakers with a broad understanding of the issues, and they set the stage for meeting this challenge by the end of the day.

The speakers urged participants to keep a broad view of innovation in their sights: They underscored that innovation in successful firms and organizations extends beyond scientific R&D (i.e., innovation includes "go to market" activity) and that innovative organizations and innovations are often rooted in entrepreneurship. They emphasized that to understand business competitiveness, studies will need to go deep into the

structure of companies and examine relationships among sub-company organizations and discover the ways in which employee skills and knowledge are leveraged by firms. They were unequivocal on the necessity of going beyond the data a company is willing to give to national statistical agencies and having a vehicle for researchers to address questions of the day in a timely, efficient fashion. No less critical was the need to engage business organizations as partners in the development of a new empirical platform for the study of innovation.

The remainder of the day therefore focused on the potential for developing a viable strategy for advancing the science and empirics of innovation. The experts were divided into three discussion groups and were first asked to pose what they saw as the key research questions in the area. After hearing presentations on new data collection modalities, the experts were asked to specify what and how information needed to be collected for researchers to begin to answer the key questions. The business representatives and computer science experts made significant contributions to these discussions, underscoring the gains from assembling a highly diverse set of experts.

Presentations on data confidentiality and data access issues introduced examples of empirical efforts with the safeguards required for a new infrastructure. These presentations substantially energized the possibilities seen by the participating social/management scientists.

Business representatives addressed the benefits from participating in an NSF-sponsored effort going forward. In large part, the sensitivities and opportunities discussed in the previous section were echoed, but with regard to willingness to participate, the sense was quite strong. One comment may be paraphrased as follows: "Our concern is not that competitors will learn about what we are doing. They already know that. We would participate to gain cutting-edge knowledge about how innovation is measured and managed in firms in other markets and industries."

All told, the experts considered innovation as both a process and an outcome; as inherently dynamic; and requiring detailed data and analysis for further understanding. Their research agenda aimed primarily at *measuring innovation* via developing a broad understanding of innovation's (1) processes and lifecycles in individual businesses, and (2) role in the global business environment and national economy. Understanding the *drivers and consequences of innovation* for individual organizations and the national economy were considered the ultimate goals.

Measuring innovation in organizations

What is innovation and where do we look to find it? Experts considered innovation broadly and mainly within a business context:

- No matter how innovation was typed or classified (product, process, organizational; disruptive, breakthrough, derivative/incremental), innovation was

viewed as a process whereby organizations put something new (research results, ideas, designs, employee knowledge) to commercial use or financial gain.[8]

- The study of innovation must involve a unit (or units) of observation applicable both *within and across* organizations, and preferably scalable.
- Data on innovation inputs and the *business outcomes* expected and/or actualized from those inputs are the major gaps in our current data system.

How do we populate these data gaps? Although the overwhelming approach was to think where to go within a typical organization to collect the needed data, the experts also emphasized the need to have scalability of results to levels relevant for macroeconomic and science policy analysis.

Unit of analysis

What is the most fruitful level of analysis for new scientific research on innovation? After considering innovation measurement at multiple levels, the discussion converged on the utility of a *project-based unit of analysis* within a given business organization (or proto-organization where innovation occurs). The possibility of filling the data gaps inhibiting transformative research on innovation were seen as especially promising at this level of analysis.

The project is the basic unit of production in many services firms and the basic "unit of innovation" in many others. Although data on innovation projects may be obtained from company records for certain organizations, determining and developing the "unit of innovation" itself is an important subject of research for many others. Characteristics of this unit of innovation include the following:

- The unit should capture the entire lifecycle of an innovation (or the expected lifecycle).
- Depending on the precise research question, the unit should be scalable upwards (firms, groups or networks of firms) or downwards (teams, social networks, entrepreneurs).
- The unit must be associated with an outcome that determines the degree of success of the research project/initiative/idea.

For a real world management process that would appear to have many of these characteristics, see the description of what Xerox calls "managing value chains" in the report of the National Academies' Board on Science, Technology, and Economic Policy, "Research and Development Data Needs: Proceedings of a Workshop," pp. 17-19.[9]

A project-level unit of analysis provides natural "scope advantages," of which experts noted the following as examples:

- Conducting innovation research at the project-level captures the development of customized services and creative solutions to general problems, areas beyond the scope of existing studies whose focus is scientific R&D yet especially relevant to

[8] This view is consistent with the definition used in the report to the Secretary of Commerce by the Advisory Committee on Measuring Innovation in the 21st Century Economy.

[9] Remarks by Charles Duke, Xerox Corporation (Washington, D.C.: The National Academies Press, 2005).

business strategists (understanding the complete value chain) and economic analysts (understanding the service economy in the United States).
- The project-based unit of analysis is especially relevant for cutting across the multiple organizations (alliances, universities) that play important roles in innovations with long lifecycles and whose processes are complex (e.g., "open" innovation processes).

Owing to the multinational nature of many businesses, global considerations increasingly enter business strategic decisions, and innovative activity of U.S.-headquartered firms is not necessarily located in the United States. As a result, many research questions need data relevant for the multinational, global domain.

- The project-level unit of analysis is amenable to the collection of innovation data across and within national boundaries for multinational firms.
- In work in this area, the results of the "community innovation surveys" conducted in many OECD countries, as well as BEA's surveys of multinational companies, and other sources of global business information, were considered important complements of the new infrastructure.

Innovation process microdata

What data are needed to determine the economic and social value created by innovation in organizations? What are the characteristics of successful innovations?

- To determine the economic value created through innovation, data on (1) the full costs of an innovation project over its lifecycle and (2) a measure or measures of the outcome of the project (preferably one in a dollar metric) are needed.
- In general, detailed data on workers—their skills, their responsibilities, and their knowledge—including their flows across companies were desired for transformative research on the combined process of entrepreneurship and innovation.
- Data on the social and cultural aspects/determinants of innovation were also desired, especially for exploring the emerging area of social networks.

The creation of basic data on *innovation net outcomes* allows the study of the determinants and characteristics of innovation successes and failures. The roles of: organizational practices (employment and management); organizational characteristics (employee knowledge and skills, business model, IT use); environmental and cultural factors (location and networks); entrepreneurial factors (firm age and origin); as well as other factors (dynamics) can be examined in terms of degree of success.

Existing studies have associated many of the above-mentioned factors with firm-level market valuations and/or labor productivity. But the established associations generally are not structural. How *do* firms appropriate the knowledge of their employees? How *do* enhancements the work environment promote innovation? Creating a new data infrastructure opens richer and deeper opportunities for exploring these questions.

Much of the economic and social data called for are relatively basic because, as previously noted, some companies currently keep records of costs and margins along

project lines. For other companies, the underlying production and innovation processes in participating companies will need to be identified and the project-level data on units of innovation created accordingly. And, where innovations are the outcome of a diffuse, creative and risky process with a long time lag between spending and payoff, specifying and determining the basic data and the role of expectations will be challenging.

The approach would be to determine a selected group of firms willing to participate in transformative research on innovation processes in organizations. Ideally, the firms in the group would represent a wide variety of industries. Approaching the collection of project-level innovation data using a "census approach" was considered impractical and too costly.

To pioneer the needed data creation/collection effort, an interdisciplinary team of sponsored social and computer science researchers will need to work with strategists and research-oriented practitioners from participating companies to determine specific research questions and collect the data needed to address them. The initial results will likely be heterogeneous across companies and industries. But as the evidence emerges, it will begin to yield salient insights on the characteristics of innovation successes and failures. Accordingly, participating companies may be among the first to make practical use of the new research results in formulating strategy.

Business function microdata

Research data on innovation processes from selected companies are unlikely to be representative of innovation inputs and outcomes in the economy as a whole. What can be done to preserve the scientific inquiry?

The collection of representative data by business function/process is a necessary component of a data infrastructure for the study of innovation and organizations. Business function concepts have been found to have substantial meaning and applicability for respondents to business surveys, largely because the concepts are grounded in the popular value chain model of firm activities introduced by Michael Porter in his 1985 best-selling book, *Competitive Advantage* (New York: Free Press).

Business processes and business functions include procurement, operations, products and services development, and the like. The approach would be to sample the universe of employer firms and collect selected economic data (e.g., total spending and employment costs) by business function/process. The Mass Layoff Statistics program of the Bureau of Labor Statistics has experimented with the collection of data by business function and finds that most establishments define their activities in terms of business functions.[10]

[10] Sharon Brown, "Business Processes and Functions: A New Way of Looking at Employment," *Monthly Labor Review* (forthcoming). See also Timothy J. Sturgeon *et al*., *Why We Can't Measure the Economic Effects of services Offshoring: The Data Gaps and How to Fill Them*, MIT Industrial Performance Center Services Offshoring Working Group Final Report, 2006; and Arie Y. Lewin, "About the International Offshoring Research Network (ORN) Project," mimeo, Duke University, Fuqua School, July 2008.

The linking of innovation project-level data with business function/process-level data would yield a new microdata laboratory for studying innovation and organizations. The inclusion of the business function/process-level survey data uniquely provides:

- A rich longitudinal resource for standardizing and benchmarking the data and findings from the project-level innovation process research and for linking them to other sources of data useful for studying innovation.

- The ability to design new innovation indicators; for example, national totals of business spending on "new product and/or service development" and "strategic management/business process development" are an indicator of business investments in innovation.

This research data infrastructure—the base microdata laboratory and ongoing survey apparatus—could be supported and managed in a fashion similar to the General Social Survey: Core statistics collected every one or two years; supplemental data modules added to address specific research questions and business and policy issues of the day; and wide researcher access via a system with appropriate safeguards and standards.

Measuring innovation in the national economy

What does innovation do for the national economy? What are the indicators of future innovation success and/or failure? How can gains in social welfare be fostered through innovative activity?

Gains in social welfare, living standards, and national competitiveness are driven by increases in productivity. As a result, a key recommendation of the Advisory Committee on Measuring Innovation in the 21st Century Economy in their report to the Secretary of Commerce (2008) was to strengthen the nation's statistical system to yield better measures of productivity. Although productivity change is linked to S&E (science and engineering) activity historically, a new consensus links the underlying growth of national productivity to innovation in organizations and knowledge creation more broadly.

Because broad indicators of innovative activity are next to nonexistent, policy analysts still rely on S&E indicators—data on patents, R&D inputs, the S&E workforce, and the like as gauges of inventive/innovative activity. Activities of the modern business organization, such as market research, "soft" design and development, the creation of entertainment and artistic originals, and investments firms make in training employees and developing new business models and strategies—activities associated with innovation—are unmeasured and missed in the discourse.

The sources-of-growth model used by economists provides a structure for measuring the contribution of innovation inputs to productivity growth, and such inputs are thereby an innovation metric. This approach requires that the nation's economic accounts treat investments in innovation on par with those in equipment, structures, and land, however. The national accounts of the United States will be moving in this direction with the capitalization of scientific R&D early the next decade. The focus on scientific R&D is

understood given limitations of the available data. Nonetheless, the Commerce Advisory Committee urged the BEA to move towards building an "innovation account" applicable more broadly to business in the 21st century.

Measures of innovation inputs, however broad, are limited as indicators because the productivity of the inputs themselves generally is not known. This long has been a limitation of the practice of using S&E inputs to look for areas of under-investment to suggest how policy-makers should allocate resources to promote economic growth. The research data infrastructure discussed above would yield new insights, stronger empirics, and thereby a strengthened understanding of the role of innovation in economic growth. Its microdata creation is squarely aimed at measuring and determining the factors that affect the degree of innovation success in terms of the characteristics of projects and the organizations and broader environments that generate them. For example:

- Basic information on innovation projects (lifetimes and relative costs/prices, for example) can be used to improve the placing of different types of national-level innovation investments on the same footing.
- The ability to use the new microdata to further study the connections between commercial success and government-sponsored research and entrepreneurship helps policy-makers formulate strategies for advancing the rate and direction innovative activity.

Understanding the innovation environment

What is the role of innovation and technology—and their relationship to one another—in the broad, global environment? What social and environmental factors affect the rate and direction of innovation and technological development? What are the social and cultural impacts of technological innovations?

What about the impact of new technologies on innovation and the structure of organizations? Web 2.0 is coming at us: How will that change innovation possibilities and outcomes? What about the implications of "green" tech? Who will be affected by these and other new technologies (customers, suppliers, and partners)?

The experts asked many questions in this area and signaled the importance of understanding of the broad innovation environment for formulating effective business strategy and government policy. The relationship of information and communication technology (ICT) to innovation was mentioned most frequently, noting the potential for disruptive innovations (e.g., the growth of the internet) that according to Clayton Christensen upset prevailing business models or create entirely new markets. [11]

Social scientists and policy-makers regarded IT capital as an important factor of production in the modern economy; the role of IT in enhancing the productivity of investments in innovation itself (e.g., scalable business methods) also is emerging. A broader understanding of the inter-relationships between the innovation process and ICT technologies is needed for effective policy formulation. Equally important is to understand its social and cultural impacts.

[11] Clayton M. Christensen, *The Innovator's Dilemma* (Boston: Harvard Business School Press, 1997).

Leveraging research datasets

The experts at the workshop mentioned numerous existing data sources and results that could be leveraged along with the collection of new data and generation of new results.

Chief among them were the results of the NBER's patent database and its international counterpart the OECD REGPAT database maintained by the OECD[12]; measures of firm performance from corporate reports (soon to be available in XBRL from the SEC)[13]; and available models and databases compiled as part of publicly-funded projects, including the Patton and Kenney IPO database, Zucker and Darby's STARS database, and the FIVE project data among others.[14] Including the available information on financing start-ups via venture capital also was mentioned.

An interesting suggestion—perhaps sparked by the novelty of the methods that were reviewed in the presentations at the workshop—was to create a web-based bulletin board to foster a community whose further suggestions could be harvested in the effort going forward.

[12] See I. M. Cockburn, B. H. Hall, and M. Trajtenberg, "National Bureau of Economic Research Patent Database: Data Overview," paper presented at the *2007 Kauffman Symposium on Entrepreneurship and Innovation Data*, and S. Maraut *et al*., "The OECD REGPAT Database: A Presentation," OECD Science, Technology and Industry Working Papers, 2008/2, OECD Publishing. doi:10.1787/241437144144

[13] See http://www.sec.gov/spotlight/xbrl/xbrlwebapp.shtml for further information.

[14] D. Patton and M. Kenney, "Initial Public Offering Database: Presentation." 2007 Kauffman Symposium on Entrepreneurship and Innovation Data. Available at: http://ssrn.com/abstract=1028004; L.G. Zucker and M.R. Darby, "Star Scientists, Innovation, and Regional and National Immigration," NBER working paper #13547 (October 2007) and references to previous work therein; and C.E. Helfat and S. Klepper, "Firm and Industry Evolution and Entrepreneurship (FIVE Project): Data Overview." 2007 Kauffman Symposium on Entrepreneurship and Innovation Data. Available at: http://papers.ssrn.com/paper=1028022.

Acknowledgments

The recommendations for new strategic directions and ways forward were based on the presentations and many excellent suggestions of participants in the workshop. Additional thanks go to participants Jeanette Blomberg, Martin Fleming, and Charles Hulten who provided comments on a preliminary draft.

Special thanks go to Julia Lane, Program Director of the Science of Science and Innovation Policy (SciSIP) program in the Social, Behavioral, and Economic (SBE) Directorate of the NSF. Julia oversaw the shaping of the agenda for the workshop and provided contacts to researchers in the NSF's CyberTrust program. Program Directors Jack Mazeros and Laurel Smith-Doerr also contributed and ensured, along with Julia, that a broad range of social science researchers participated in the event. All three provided comments on a preliminary draft of this report.

Colleagues Bart van Ark and Linda Barrington of The Conference Board worked with me to develop and manage all aspects of the workshop, including the preparation of this report. Conference Board President Gail Fosler supported our effort from its earliest stages to the completion of the report.