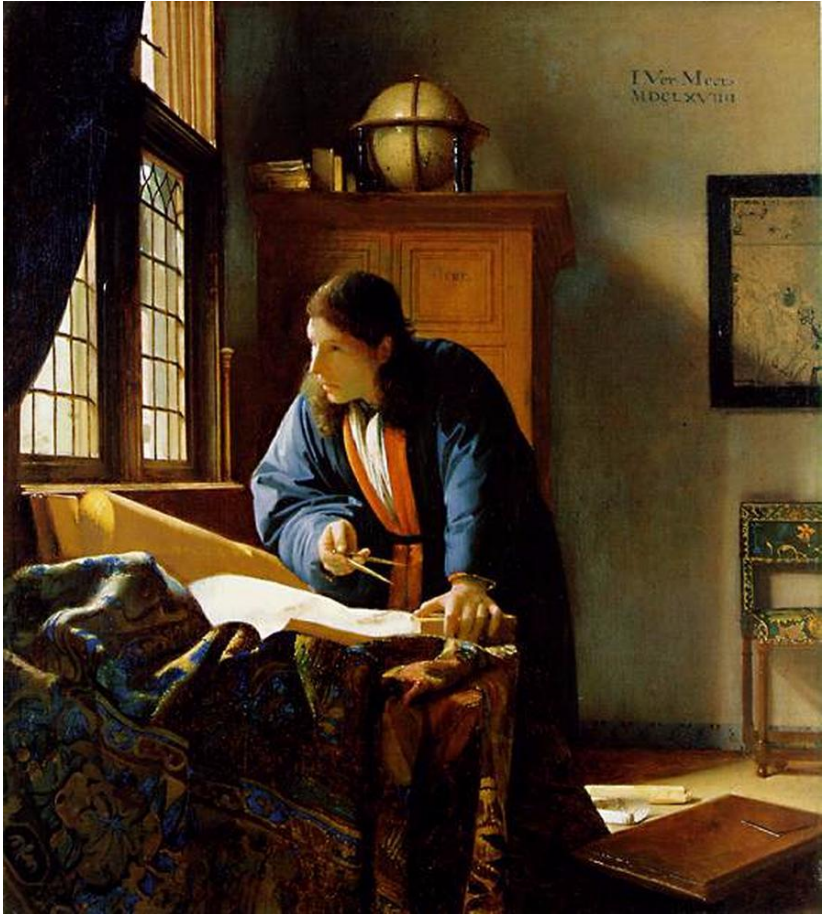


INNOVATIVE DATA SOURCES FOR REGIONAL ECONOMIC ANALYSIS



Maryann Feldman, Alex Graddy Reed, Lauren Lanahan,
Glenn McLaurin, Kari Nelson, and Andrew Reamer

INNOVATIVE DATA SOURCES FOR ECONOMIC ANALYSIS

Maryann Feldman

Alex Graddy-Reed

Lauren Lanahan

Glenn McLaurin

Kari Nelson

Andrew Reamer

Acknowledgements: Funding for this project was provided by the Ewing Marion Kauffman Foundation; the Economic Development Administration, U.S. Department of Commerce; the Alfred P. Sloan Foundation; the Lumina Foundation; and the Council for Community and Economic Research.

The front image is Johann Vermeer's *Geographer* (1668-9), Städelsches Kunstinstitut, Frankfurt am Main

INNOVATIVE DATA SOURCES FOR ECONOMIC ANALYSIS

“Developing better data is part of the Kauffman Foundation's long-term strategy for advancing better research and policy on entrepreneurship and innovation.”

E.J. Reedy (2012)
Kauffman Foundation

“I would like to make a modest proposal – that we leverage private sources of economic data to improve our statistical infrastructure.”

Alan B. Krueger (2009)
Assistant Secretary for Economic Policy
U.S. Department of Treasury

Table of Contents

The Data Imperative	1
The Opportunity of Innovative Data Sources	5
Filling Knowledge Gaps with New Sources of Data.....	8
Reflective Conclusion	16
Data Fair Executive and Evaluation.....	17
Conference Development	18
Conference Exhibitors and Participants	20
Conference Activities	22
Conference Assessment, Impacts, and Outcomes	27
Suggested Improvements.....	31
Symposium: Use of Innovative Data Sets for Regional Economic Research	33
Looking Ahead	35
Potential Utility for an EDA Performance Evaluation System	37
Third-Party Microdata Sources	38
Regional Economic Database.....	47
Data Tools	55
Innovative Data Sources.....	57
Altmetrics Total-Impact	59
Amazon Web Services	61
American Community Survey	63
Business Dynamics Statistics	65
Business Employment Dynamics	67
Business R&D and Innovation Survey	69
Burning Glass: Labor/Insight	72
CB Insights	74

Circling the Research Triangle.....	77
Cost of Living Adjustments for the New Supplemental Poverty Measures.....	79
Data.gov	81
Dataverse Network Project.....	83
DataWeb for the DataFerrett	85
Discovery Logic	87
EMSI Analyst.....	89
Export Nation 2012	91
Factual	93
GeoIQ	95
Green Goods and Services Survey	97
Innovation in American Regions.....	99
Kenny-Patton IPO Database.....	103
Local IDEAs (Indicator Database for Economic Analysis)	105
Longitudinal Business Database	107
Local Employment Dynamics	109
Mendeley	113
Monster Real-Time Labor Intelligence	115
Moody's Analytics	117
National Establishment Time-Series (NETS)	119
National Student Clearinghouse.....	121
O*NET Data Collection Program	123
PriceStats	125
Research Data Centers	127
Regional Price Parities	129
Rural Establishment Innovation Survey.....	131
S&E Indicators State Data Tool	133
S&P Capital IQ.....	135
Science of Science (Sci ²) Tool	137
SciENCv: Science Experts Network and CV	139
SciVal, Elsevier	141
STAR METRICS	143
Statewide Longitudinal Data Systems.....	147

STATS America	149
STATT: Statistics Access for Tech Transfer.....	151
TEN: The Evidence Network	153
ThomasNet's Product News Room	155
U.S. Cluster Mapping Project.....	157
USAspending.GOV	159
USPTO: U.S. Patent and Trademark Office.....	161
University Economic Impact Metrics.....	163
Wanted Analytics	165
Web of Knowledge.....	167
Windows Azure Marketplace—Microsoft	169
YourEconomy.org	171
Conference Attendees	173
Symposium: Use of Innovative Data Academic Conference	206
Common Abbreviations	209
Common Definitions of Terms	211

The Data Imperative

Increasingly, there is recognition that the future of the nation's prosperity is substantially determined by activities at the local, regional or sub-national level. Firms are simply more innovative and more productive when they can easily access capacities that support and encourage the creation and utilization of knowledge, promote creativity and risk-taking entrepreneurship and business development, provide a skilled workforce, offer access to capital and provide a well maintained physical infrastructure. More specifically, the ability of U.S.-based firms to sell goods and services to the nation and the world is very much a function of their regional context.

U.S. regions are increasingly vulnerable to the increasingly sophisticated economic capabilities of developing nations. Consequently, an important aspect of federal economic policy has become investing in regional economic capabilities that cultivate the ability of firms to compete on the world stage. There is a need for federal, state and local policymakers to understand and respond to the rapidly evolving geography of U.S. economic activity.

Information is among the most essential and cost-effective of the policy tools at government's disposal. Both policymakers and market participants require current, accurate, detailed economic statistics to monitor, assess, and respond to competitiveness issues and opportunities. Addressing these needs is a relatively low-cost endeavor that increases the function of markets and highlights both opportunities and potential problems.

Despite the nation's competitive vulnerabilities, federal economic policy and data collection has remained tied to traditional business cycle management tools with a focus on data useful for fiscal and monetary policy. As a consequence, there has been less recognition of the need for coherent, evidence-based regional policy.

In particular, the federal government has not addressed the value and role of national and regional economic statistics. Moreover, federal economic statistical agencies have not traditionally viewed the provision of regional statistics that understand industrial competitiveness as central to their mission. Consequently, while federal regional statistics programs are certainly underfunded there is a

need for additional products that are better matched to user needs.

An increased policy emphasis on innovation, entrepreneurship and emerging industries are the building blocks of our 21st century economy. Given that these are relatively new policy imperatives, these concepts were not addressed when legacy systems were designed. Once again, traditional economic statistics have little to offer either the public or private sector investor.

A plethora of regional innovative data sources and tools are emerging, spanning federal statistical and mission agencies, commercial firms, universities, and nonprofit research organizations. These data sources have the potential to greatly increase the abilities of regional businesses, governments and advocacy organizations to understand and respond to issues in and opportunities for economic competitiveness in a timely manner. In addition, these data provide the resources for researchers to understand the nature of regional economies, the factors that influence competitiveness, and principles for the design of national and regional policies and programs to promote economic development, growth and prosperity.

This report's aims are to provide an overview of innovative sources of regional data. The next chapter establishes the promise of new data sources to inform policy and research to guide policy. Chapter three provides assessment of the potential utility of innovative economic data sources in an EDA performance evaluation system. To vet our approach, a data fair was held at George Washington University on May 7 & 8, 2012. This event is described and evaluated in the fourth chapter. We then provide a 2-3-page overview for each of the data sources that were included in the data fair and that we have identified as potentially useful for regional economic analysis. These sources are listed alphabetically. This material was core of the briefing book provided at the event. The contact information for all of the presenters is included with each source. We then provide a list of the attendees followed by three sections on the Symposium that follow the fair, Common Abbreviations, and Common Definitions, respectively.

The Opportunity of Innovative Data Sources

To better study innovation requires new approaches and innovative data sources.

Recent advances in information technology provide an unprecedented opportunity to collect, organize, analyze, disseminate, and visualize large volumes of data generated from private and public administrative records. Concurrently, new statistical methods make possible the creation of microdatabases that allow new ways of studying economic behaviors while, when necessary, fully protecting confidentiality. Further contributing to new data collection efforts is an increased policy emphasis on innovation, entrepreneurship, clean tech, and other such building blocks of our 21st century economy.

As a result of these advances, a plethora of innovative data sources and tools is emerging from a number of organizations from federal statistical and mission agencies, commercial firms, universities, and nonprofit research organizations. These data sources have the potential to transform our understanding of the phenomena that provide the basis for our economic well-being. Improved understanding should lead to better designed, more

effective policies and programs for stimulating economic growth and development.

Of particular importance in observing economic growth and related human activity are the nation's regions. Recognized around the world as providing the foundation for economic growth, regions are organically organized, relationship-laden geographies tied by common economic interests and catalytic infrastructures. Understanding how regional economies work and how they could work better will boost the likelihood of a bright economic future for the nation and its residents.

A regional economy is the sum of transactions among firms and people. Importantly, these transactions are not confined within political or Census boundaries. In this time of economic volatility and vulnerability to global competition, static tabular data products that aggregate the number of businesses, jobs, and workers by artificially imposed, static jurisdictions rather than by current, more flexible, organically occurring economic communities are not sufficient for generating an understanding that leads to intelligent public and private economic investments.

Innovative sources of economic data, however, offer analysts new ways of understanding regional economic activity. This is made possible through the analysis of the following: the actions of

individual institutions, entrepreneurs, investors, and workers over time and actual space; the relationships among these actors; and the outcomes of these relationships. Consequently, these data sources have far greater capacity to describe how regional economies work than more traditional data products. These contemporary sources, in turn, facilitate better informed private and public actions to create jobs, income, and profits.

Our objectives are threefold: (i) to raise the awareness of policymakers, practitioners, and researchers about innovative data sources useful for regional economic development analysis and policy; (ii) to build a community of interest on the use of these new sources; and, ultimately, (iii) to advance the availability and reliability of useful regional economic data.

This document considers the potential for innovative data sources to enhance research, practice, and policy making. Our hope is that awareness of these data sources will stimulate conversations among public and private data providers and economic analysts. Through these efforts we hope that a growing community of common interest may form to invest in the new generation of regional economic data and advocate for greater use of data to inform policy.

Filling Knowledge Gaps with New Sources of Data

Motivating this e-book is a central question for policymakers and researchers: Why do some regions thrive and prosper while others stagnate or fall into relative decline?

At the moment, the data needed to answer this question are unavailable or unknown to researchers and analysts. Legacy federal economic data products do not meet regional data user needs because the traditional mission of federal economic statistical agencies has been to support federal macroeconomic policy and to guide the distribution of federal funds to political jurisdictions with particular economic characteristics, such as high unemployment.

While entrepreneurship and innovation are frequently offered as policy panaceas for all that ails the U.S. economy, we observe many cases in which regions do not benefit from public and private investments to promote these efforts and so fail to adapt to changing economic challenges. In no small part, the failure of these public investment strategies is due to a dearth of information. Analysts lack the data to accurately diagnose their regions' economic problems, and policymakers lack the tested principles to act on suggested diagnoses. In the absence of useful data and information,

policymakers are often left to mimetically adopt programs and policies that are said to have worked in other places. However, policymakers tend to mimic policy actions in other regions with quite different characteristics. Until analysts have access to better data and information, policymakers will continue to make decisions based more on anecdote, intuition and hope than on empirically sound evidence.

Most importantly, the communities most relevant for regional analysis do not fit neatly within predetermined boundaries. Areas such as Silicon Valley, Route 128, and the NC Research Triangle – the archetypes of successful regional economies – have complex geographic shapes that have evolved organically. These regional economies are built on the location of prominent institutions and firms, are influenced by existing transportation routes and land use patterns, and expand out of seemingly idiosyncratic and serendipitous events. Their spatial patterns do not acknowledge political jurisdictions or census geography, but instead follow a logic that motivates firms to locate near others with similar products, markets, and employees with the requisite workforce skills.

Not only do regional economies blithely ignore town and county boundaries, but they disregard state boundaries as well. Furthermore, the

geographic clustering of innovative, creative firms in small places such as multi-tenant buildings, neighborhoods, or adjacent industrial parks is often invisible when data are available only for larger, rigidly defined political units, such as counties. Data that suggest a cluster at the county level may indeed mask several geographically, and often technologically, distinct clusters grounded in distinct social relationships and that operate based on unique needs, production logics and positions in the value chain. The use of aggregate data often leaves promising early-stage activity overlooked.

Locating market actors across real economic space requires access to digital data that facilitate flexible, user-determined analysis of salient characteristics. Such data may be easily drawn from standard records, scraped from the web, pulled from voluminous documents through text analysis, found on an open data platform, purchased from a third party, and integrated with other digital data sets. More data is available than ever before. It is now possible for researchers to link and analyze microdata to explore economic relationships within unique self-defined economic boundaries.

The innovative firms driving regional economies are themselves fluid and difficult to classify. As firms struggle to survive, they often modify their products or services, but there is no time or

incentive for them to update their NAICS industrial classifications—the main mechanism primarily used for understanding industrial activity.

For purposes of understanding new activity, detail in patent documents provides an idea of where a company is headed with their inventive activity, while new product announcements offer a mechanism to understand where firms are placing their bets in the market. Both patents and product announcements thus provide information on the economic future of firms. Understanding forward-moving industrial activity requires classification schema to be fluid and malleable, perhaps based on text-mining, algorithmic programming to define relational attributes. Static classification schemes will never provide an understanding of the emerging technologies that have the greatest promise for building new industries and setting regions on a new growth trajectory.

The time lag inherent in the collection of traditional data sources means that our analysis is consistently retrospective, lagging the current reality that we are trying to analyze. These limitations reflect the technologies of a prior time, before we had Internet access to current administrative data and the computing power to manipulate them at our fingertips.

Traditional datasets and industrial classifications do not describe the relationships between organizations across space; however, innovative data sources allow the identification of networks and social relationships among firms, between firms and institutions, and between their connections both inside and outside the region. This enables analysts to determine points of leverage for economic expansion. These relationships are particularly relevant in the form of institutions such as universities, trade associations, business services, and other quasi-government entities that are important to innovative activity and provide the foundation for economic vitality in regions.

Long-standing forms of economic data also impede analysts' ability to view the behavior of market actors over time. Analysis of regional innovation systems, for instance, has been restricted to looking at a series of disconnected snapshots, which can easily lead to inappropriate or incomplete conclusions that ignore the complexity of these endogenously and historically path-dependent systems. Though case study narratives provide insights, they lack both analytical power and generalizability. However, using advanced IT, researchers now can construct longitudinal microdata with which they can follow the dynamics of emerging and mature industries, understand the theoretically important links between firms and

institutions, and measure the employment outcomes of different approaches to education.

In the end, efforts to build and maintain regional economic advantage often involve some combination of bottom-up efforts by regional public and private actors and top-down resources invested by the federal and state governments. These efforts are far more “hit and miss” than would be the case if actors were able to access significantly improved data and information. Innovative data sources—based on advanced IT, new statistical methods, and untraditional research topics—offer the opportunity to fill this knowledge gap. Realizing this opportunity depends on the efforts of research organizations, federal decision makers, philanthropic foundations, entrepreneurs and private investors, and economic and workforce development organizations to do the following: to support, demand, and create a market for these innovative data sources; to continually redefine the state of the art; and to bring to bear the degree of creativity, risk-taking, and entrepreneurship being asked of regional economic actors.

A central aim of the conference was to serve as a springboard for such action and to:

- Provide access to economic data not otherwise available (such as the nascent

effort by the Association of Public and Land Grant Universities to measure the economic impacts of universities)

- Provide user access to microdatabases, allowing analysis of individual records (often of a universe of economic actors for a specific location) that may be dictated by the user (for example, University of North Carolina's Circling the Research Triangle Project)
- Allow analysts to see economic activity in real time (for instance, through the analysis of on-line job advertisements by WANTED Analytics, Burning Glass, and Monster Government Solutions)
- Enable analysts to study economic dynamics over space and time (for example, the Census Bureau's Local Employment Dynamics Program)
- Offer new ways to visualize and analyze data (as do GeoIQ and Indiana University's Science of Science [Sci²] Tool)
- Add value to existing datasets (for example, EMSI, Harvard's Cluster Mapping Project, and Indiana University's Innovation in American Regions web tool)

- Provide access to, integrate, and facilitate applications development for datasets from multiple sources (of which there were many examples at the conference; including Amazon Web Services, Moody's Analytics, Windows Azure Marketplace, Factual, TheDataWeb, and Data.gov)

In addition, the conference provided a venue for knowledge exchange, collaboration, and contractual relations among these various types of providers in attendance. Consequently, they decided to expand the number of presentations and were overwhelmed by the positive response.

It seemed that an open invitation to bring together interested parties would better stimulate new relationships among both users and providers, leading to greater improvements in data availability, usefulness, reliability, and accessibility. A diverse set of users would better inform data providers about the markets for their products and services and about how those products and services could best serve user needs.

Consequently, the organizers decided to open registration and re-label the workshop as a conference. The response from data stakeholders was substantial, and the conference was about four times the size initially proposed.

A much bigger conference requires much more money. The organizers greatly appreciate the additional financial support provided by the U.S. Economic Development Administration, the Alfred P. Sloan Foundation, the Lumina Foundation, and the Council for Community and Economic Research. Their assistance allowed the event to meet the demand of those interested and in attendance and therefore to take place in its expanded form.

Reflective Conclusion

The availability of economic data and the tools to analyze them has never been greater. At the start of Feldman's and Reamer's careers, records were stored on magnetic tapes, computer programs ran from punchcards, and data were copied from large volumes lined up by year on the shelves of Federal Depository Libraries. Quantum advances now enable previously inconceivable analyses and insights, and offer the opportunity for greater and more sophisticated understanding that will be important for the policy and research required to advance economic development and growth.

Data Fair Executive and Evaluation

During the two-day conference, over 250 participants had the opportunity to visit 52 exhibits of innovative data sources useful for regional economic analysis. These exhibits were presented by public, academic, nonprofit, and commercial organizations. As noted earlier, the conference's primary aim was to raise the awareness of economic policymakers, practitioners, and researchers about innovative private and public data sources and tools useful for regional economic development analysis and policymaking. Its second aim was to offer data-providing organizations the opportunity to productively interact with users and with each other.

The principal investigators considered a data source to be "innovative" if it relied on advanced information technology (e.g., to process very large volumes of microdata), used recently developed statistical methods (e.g., synthetic data), or focused on a topic of emerging policy interest (e.g., green jobs).

This chapter describes the conference's development, exhibitors and participants, and activities. It also includes assessments and outcomes of and suggested improvements in the conference, as provided by exhibitors and participants through two post-conference surveys.

In addition, this report briefly describes a GW-funded symposium on May 9, 2012, “The Use of Innovative Data Sets for Regional Economic Research,” at which researchers made 14 presentations on their analyses of innovative data sources. The chapter ends with thoughts about future activities that build on the May 2012 conference and symposium and the potential value of such efforts to the Economic Development Administration.

The purpose of the conference, then, was to provide a forum, which offered substantial numbers of data users and innovative data source providers multiple opportunities to learn from and develop relationships with one another. The principal investigators’ belief was that these new relationships would lead to the greater use of and improvements in innovative data sources thereby leading to the benefit of the field of economic development and the nation’s economy.

Conference Development

The original funding for the conference (approximately \$50,000) was provided by the Kauffman Foundation. The proposal to Kauffman assumed conference attendance of 65-75 people, including 16-20 representatives of innovative data sources. Eventually, however, over 350 people

registered for the conference (260 participants and 90 representatives of 52 data sources) – five times the original estimate.

Conference participation grew well beyond expectations for several reasons. First, the principal investigators identified far more innovative data sources than they originally expected. They also found that innovative data providers have much to learn from each other in terms of data retrieval, manipulation, dissemination, and uses. Additionally, federal statistical agencies would benefit from exposure to innovative external datasets and tools that stimulate their thinking about how best to fulfill their respective missions.

Second, the principal investigators chose an open registration process rather than by invitation only. They thought that a by-invitation-only approach would reduce the conference's value because participants would be limited to users known to the principal investigators and likely to each other. Conversely, they believed that open registration would bring together many people who did not know each other and so would offer a far greater opportunity to create new relationships among users and providers that could lead to improvements in data availability, usefulness, reliability, and accessibility. The principal investigators also thought that a diverse set of users would better

inform data providers about the markets for their products and services and how those products and services could best serve user needs.

Third, conference registration was free, which allowed many participants to attend who otherwise would not have been able to (particularly federal statistical agency staff).

The unplanned growth in the number of conference participants resulted in the project budget more than doubling. As a result, the principal investigators sought funding from additional sources and were successful in that effort. Contributions from the Alfred P. Sloan Foundation, the Lumina Foundation, and the Council for Community and Economic Research allowed George Washington University to cover all project costs.

Conference Exhibitors and Participants

As noted, 52 innovative data source exhibitors had a table at the conference. Each of these is reviewed in the following section. Distribution by type of organization was as follows:

- Federal statistical agencies (19)
- Academic institutions (9)
- Commercial vendors (20)
- Membership organizations (2)
- Think tanks and research organizations (2)

While most exhibitors were present both days, a small number were on the floor for only one day due to other commitments.

To aid conference attendees, the principal investigators prepared a briefing book (an early iteration of this document) and a web site that categorized exhibitors among topics and tools:

- Topics
 - Business Creation and Development (9)
 - Jobs, Workforce, Education & Labor Markets (10)
 - Longitudinal Databases (6)
 - Networks & Relationships (5)
 - Prices & Costs (4)
 - Regional Industries & Economies (15)
 - R & D, Innovation, & Commercialization (18)
- Tools
 - Big Data, Open Data Platforms, & Web Services (6)
 - Data Intermediaries & Integrators (3)
 - Data Analysis & Visualization Tools (4)

Some exhibitors were listed in more than one category. The web site provided a direct link to each exhibitor, as well as a number of innovative data sources not represented at the conference.

The 260 registrants represented a diverse array of organizations, including:

- Nonprofit organizations (25 percent)
- Universities (21 percent)
- Federal policy/program agencies (20 percent)
- Federal statistical agencies (13 percent)
- State and local government agencies, economic and workforce development organizations, membership associations, consulting firms, and other organizations (21 percent)

A list of participants, in alphabetical order by organization, is provided in the section *Conference Attendees*.

Conference Activities

The two-day conference was held in a “data fair” format in the 3,900 square-foot Grand Ballroom of the Marvin Center at George Washington University.

The principal investigators’ work plan for conference structure was as follows:

- Categorize each exhibitor by one or more topics, along the lines of the categories listed above.

- As part of the registration process, ask each participant to indicate the top three categories of interest.
- According to the identified categories of interest, assign the participant to three small topic-specific groups (8-10 people each), one for each of the first three half-days of the conference. Registered participants were distributed among 30 small groups for the morning of May 7, 32 small groups for the afternoon of May 7, and 32 small groups for the morning of May 8.
- At the beginning of each half-day, ask each small group to explore the exhibits covering the respective topic and then reconvene to assess what the group had seen, posting group comments on a live conference blog and individual assessments on a confidential on-line evaluation form for each exhibitor.
- Also in each of the first three half-days, the group sessions were to be briefly interrupted with a short talk by a peer referee in the field of regional economic data, including Robert Groves, Director, U.S. Census Bureau; Anthony Carnevale, Director, Georgetown University Center on Education and the

Workforce; and Mark Doms, Chief Economist, U.S. Department of Commerce.

- Keep the fourth half-day (Tuesday afternoon) open, with no groups to allow for unstructured discussions.
- After the conference, organize and synthesize blogs and exhibitor evaluations into conference proceedings.

After the welcome and introductory remarks, participants were asked to go to their assigned group meeting place, get organized as a group, and then proceed to the exhibit floor. However, it quickly became clear that without a trained facilitator for each group, the process was insufficiently structured to work as originally planned. Many people simply went to the floor as individuals, skipping the group format. Among the groups that met, a number found the discussions so interesting that they found it difficult to stop and go between all the exhibitors on the floor. When some groups came to the floor, they quickly broke up, as individuals felt drawn to different exhibits. Also, some people saw acquaintances and stepped aside to chat. Very few groups blogged about their experience and very few individuals posted exhibitor evaluations.

As a result, at the beginning of the Monday afternoon session, the principal investigators announced that the small groups would not meet during the remainder of the conference—participants were told they were free to visit exhibits as they wished. Throughout the first day and the morning of the second, the floor was quite crowded and bustling as participants moved among exhibitors, participants spoke with participants, and exhibitors had the opportunity to learn about other data sources. The crowd thinned out considerably on the afternoon of May 8, as by then most participants saw the exhibits of particular interest to them.

As planned, Anthony Carnevale spoke halfway through the May 7 morning session. The themes of Dr. Carnevale's remarks included: decisions by labor market participants were more frequent and complex than previously because of greater labor market volatility, which increased emphasis on attaining a postsecondary credential; that current, reliable, detailed regional statistics are essential for intelligent decisions by employers, educators, workers, students, public purpose organizations (such as workforce boards), and governments at all levels; that advances in information technology are enabling the development of an array of new, useful data sources; and insufficient federal funding for socioeconomic statistics is undercutting labor

market participants' capacity for intelligent decision-making.

Census Bureau Director Robert Groves spoke on the afternoon of May 7. He noted the stagnant budgets and increasing public antipathy towards federal statistical surveys and, simultaneously, the ongoing emergence of a considerable number of administrative record-based Big Data resources in both the public and private sectors. He then indicated the necessity (from a budgetary perspective) and the opportunity (from a technology perspective) for federal statistical agencies to increasingly rely on administrative record-based data and for these agencies to be creative and collaborative in this effort.

Commerce Department Chief Economist Mark Doms, the scheduled speaker for the morning of May 8, was unable to attend.

Food and drinks were made available throughout the conference, which encouraged attendees to stay in the conference area and continue to meet with others. Aiding this process was direct access from the conference room to an outdoor rooftop deck with picnic tables.

Conference Assessment, Impacts, and Outcomes

During and after the event, a substantial number of participants and exhibitors spontaneously told the principal investigators that they had a very positive conference experience. They said they appreciated the opportunity to learn about new data sources and new data-related technologies and methodologies. Additionally, they valued the opportunity to make new professional connections. Representatives from a number of federal statistical agencies and nonprofit and commercial data providers said they were pleased to learn about, and impressed with, innovative efforts taking place in other sectors. The principal investigators are aware of several instances in which a federal agency and a commercial data provider agreed to explore collaborative possibilities.

The large majority of attendees spoken with requested that the principal investigators convene another, similar conference in the near future – perhaps on an annual or bi-annual basis.

Immediately after the conference's conclusion, the principal investigators emailed attendees an evaluation survey. Survey results include the following:

- 88 percent of respondents would attend a similar data fair in the future (n=41)
- 86 percent of respondents found the data fair very informative (67 percent) or somewhat informative (19 percent) (n=42)

Participant respondents (that is, non-exhibitors) provided a substantial number of complementary comments about the data fair. Various respondents valued the breadth and diversity of exhibitors; the quality of the exhibitions; the number and quality of the attendees; the open format, which provided the opportunity to gather information about data sources and uses, engage in open-ended conversations, and build relationships; the opportunity for on-the-spot idea sharing and brainstorming; and the lack of a registration fee.

Participant respondents also identified specific interactions that they found of particular value. Collectively, respondents (n=23) highlighted 18 of the exhibitors. A number appreciated learning about the sophistication of the federal statistical agencies. Several liked having the opportunity to talk with senior government officials, leading academics, and students.

Participant respondents were mixed about the principal investigators' effort to have people

work in small groups. While they saw the potential of small group activity and appreciated the contacts and discussion, they thought that the principal investigators provided insufficient structure and instructions. Moreover, the groups were too large to move easily through the data fair.

Exhibitor respondents were quite positive about the event in general, particular experiences and interactions, and the difference between this event and other exhibitions. Exhibitors appreciated the diversity of attendees and presenters; the opportunity to build a customer base; the opportunity to network and explore collaborations with other exhibitors; and the opportunity to learn about competitor products and services. Compared to other exhibitions, they found that data fair provided the opportunity to reach a broader audience; have more productive interactions; and deepen connections with existing customers.

A second survey emailed one month after the event indicated that the large majority of participant and exhibitor respondents contacted or planned to contact people they met at the data fair.

- 80 percent of attendees contacted or planned to contact an exhibitor (n=35)

- 81 percent of exhibitors contacted or planned to contact attendees (n=21)
- 63 percent of exhibitors contacted or planned to contact other exhibitors regarding possible collaboration (n=19)

Immediately after the conference, and then one month out, participant respondents identified an array of expected and actual follow-up efforts. These include: visit data websites; contact exhibitors to discuss data use; use new data sources in research and to improve work products; collaborate with new contacts on research papers; enhance research proposal; create new business partnerships; share briefing book and web links with colleagues; obtain data training; teach others how to use a data tool; and attend a specialized conference.

Exhibitors' post-conference intentions and actions include: follow up with potential data users; collaborate with other data providers; make data product improvements on the basis of conversations at data fair; and adjust data tool tutorial on the same basis.

In summary, then, according to participants and exhibitors, the data fair was very successful in facilitating the following: acquainting data users with new data sources, introducing exhibitors to a broader user base and other data providers,

stimulating connections among data users, catalyzing action and the intention of action; and generating a widespread desire to attend similar events in the future.

Suggested Improvements

While respondents deemed the data fair a success overall, they offered a number of recommendations for any future event. These ideas can be categorized as follows:

- ➡ Expand the amount of space devoted to exhibits.

Several respondents thought the conference space was too small for the number of exhibits and attendees. Specific requests included more display space and wider aisles.

- ➡ Reduce the number of days devoted to the open data fair from two to one or one and a half days.

As noted earlier, attendance fell substantially after Tuesday's lunch. A number of people thought that one or one and a half days of open time would be sufficient.

- ➡ Introduce structured presentations and interactions into the event. Suggestions included:

- Data provider presentations in breakout rooms
- Analyst/researcher/academic presentations on the uses of innovative data sources, in breakout rooms
- “Lightening talks” – 15-20 minute talks proposed by individual participants on the day of the event
- Expert panel discussions in breakout rooms
- Longer talks by notable speakers (e.g., Groves, Carnevale) at set times and with the opportunity for audience engagement
- Small group discussions that are more organized. Ideas include:
 - Have a discussion facilitator for each group, with an added responsibility for seeing that a written summary of the discussion is prepared
 - Provide a written discussion guide
 - Provide each topic with its own breakout room and allow participants to self-select
 - Hold the small group discussions after giving participants several hours of open time with exhibitors
 - Have the small groups meet briefly, then disperse as individuals, then reconvene at a fixed time

- ➊ Provide participants and exhibitors with the briefing book in advance of the conference.

Due to the unexpectedly large amount of time required for preparation, the briefing book was completed and electronically distributed the evening before the conference. Consequently, a number of respondents said they were unable to make good use of it before the event (lack of time) and during the event (hard to read on smartphone and similar devices).

Symposium: Use of Innovative Data Sets for Regional Economic Research

After the principal investigators had set the conference dates and started planning, they were asked by GWIPP Director Hal Wolman to consider ways in which an academic symposium might complement the conference. Each year, the GW Provost's Office provides GWIPP with approximately \$25,000 to hold a day-long academic symposium on a public policy topic of its choice. The principal investigators proposed, and Professor Wolman accepted, a focus on the "Use of Innovative Data Sets for Regional Economic Research." Their thought was that the symposium could serve as a forum in which researchers from the U.S., Canada, and Europe would share cutting-edge work and meet like-minded peers. The symposium was scheduled

for May 9, the day following the data fair. GWIPP agreed to pay for speaker and discussant travel and hotel costs.

With the title and date set, the principal investigators identified and arranged for 16 academic researchers to provide 14 presentations organized around six topics—labor force, university R&D, companies, patents, regional industries, and regional economies. Two discussants were lined up as well, one for the morning presentations and one for those in the afternoon. Each presenter was asked to provide a short paper for distribution to others beforehand. The principal investigators also arranged for speakers who had developed their own innovative dataset to exhibit at the data fair prior to the symposium.

The symposium agenda, with speakers and presentation titles, is provided in the section *Symposium: Use of Innovative Data Academic Conference*". The sequence of topics was intended to flow from the micro (workers, firms, universities, and patents) to the regional (industries and economies) levels.

The principal investigators worked with Professor Wolman to prepare a symposium invitation list. In addition, Professor Feldman

suggested to Martin Kenney of the University of California at Davis, a speaker and the editor of *Regional Policy*, that he devote a special issue of the journal to the papers presented at the symposium. Professor Kenney readily agreed.

Attendance at the symposium, including speakers, was about 50, with a plurality from the academic community. The papers were well received, discussion was lively, and the audience learned about a number of innovative data sources, innovative analytic methods, and new findings regarding the workings of regional economies. Participants, particularly the speakers, appreciated the opportunity to connect with others with similar interests and strongly expressed the desire to hold a similar event in the near future.

Looking Ahead

In light of the success of the conference and symposium in achieving their goals and the strong interest by participants and exhibitors in having a similar future event, the principal investigators plan to seek funding to organize another conference on innovative economic data sources in the fall of 2013. Their thought is that a hiatus of 18 months provides sufficient time to allow new and improved innovative data

sources to emerge and yet is short enough to maintain a sense of community and continuity.

The Kauffman Foundation and the Alfred P. Sloan Foundation were very pleased with the results of their investment and have expressed interest in discussing possible funding of a future event.

In the meantime, the principal investigators are maintaining and regularly updating the web page of innovative data sources and are considering the idea of periodically (perhaps once a month) sending an email to conference participants and exhibitors to identify new innovative data sources and provide an update, if appropriate, regarding any future event.

It is our hope that the conference will lead to the institutionalization of a new productive mode for appreciably increasing awareness and use of innovative data sources and catalyzing ideas and relationships that can lead to further innovative data developments. The result over the long run, the principal investigators believe, will be more effective public and private sector decisions that support the economic competitiveness of the nation's regions.

Potential Utility for an EDA Performance Evaluation System

The *Innovative Data Sources for Regional Economic Analysis* conference provided an opportunity to ascertain the possible use of each of 52 data sources in the proposed EDA performance evaluation system. That proposed system has three component databases:

- Microdata on firms, universities, and other important regional development actors from external sources (commercial, nonprofit, academic, federal)
- Microdata collected from EDA grantees on grantee efforts and, as appropriate, their beneficiaries
- Regional economic conditions—aggregate data on conditions such as unemployment rate, poverty rate, average annual wage

The two microdatabases will be used to trace the outputs and outcomes of EDA investments. The regional economic database will provide context for project evaluations (ascertaining the factors that influence project outcomes) and program evaluations (comparing the relative impacts of each of EDA's programs). The regional economic

database also will be used to create a regional economic dashboard, a series of indicators that provide a sense of the economic conditions in each of the nation's regions.

This chapter identifies innovative data sources exhibited at the conference that appear to have utility as part of the third-party microdatabase and the regional economic database. It also includes several useful innovative datasets not exhibited at the conference. (Hyperlinks are provided for the latter; those for the former are elsewhere in this document.) A short third section covers innovative data tools—specifically open data platforms and visualization and analysis tools—useful for the EDA performance evaluation effort.

Third-Party Microdata Sources

With one exception, the innovative third-party microdatasets with potential value for EDA performance measurement contain information on business characteristics and activities and so are useful in assessing the impacts of EDA investments on individual firms. As such, these datasets would serve as a complement to information provided to EDA by its grantees. In the discussion below, the firm microdatasets are organized into four categories—general business dynamics, business development and entrepreneurship, R&D and innovation, and workforce.

The one exception, and in a fifth category by itself, is USASpending.gov, an online database of all federal grants and contracts.

General Business Dynamics

For the purposes of EDA performance evaluations, the Census Bureau's **Longitudinal Business Database** (LBD) is the single most valuable microdatabase. The LBD contains a multitude of establishment- and firm-specific information gathered through the quinquennial Economic Census; all annual, quarterly, and monthly Census business surveys; and various datasets of administrative records, linked over time (1976-latest year).

Access to the LBD would allow EDA to track the evolving characteristics of firms that benefit from assistance. As the contents of the LBD are confidential, one important limitation is that only Census staff and/or external researchers given special sworn Census employee status could conduct analysis for EDA performance evaluation. Discussions with Census staff indicate that creating a workable arrangement to serve EDA is feasible.

A commercial, non-confidential alternative to the LBD is the **National Employment Time Series (NETS)** database offered by Walls & Associates and built on the longitudinal linking of years of Dun

& Bradstreet business records. While the cost of NETS is substantial and the data are not as extensive and reliable as LBD data, regional researchers have found NETS to be a valuable resource.

S&P Capital IQ maintains a very large, detailed database of tens of thousands of public and private U.S. firms. While this database is less comprehensive than the others, it does concentrate on those firms that are more likely to be important regional economic engines. Information available by firm includes financials, debt capital, equity capital, key developments, and corporate intelligence. Integration of these data into the LBD or NETS would be quite valuable.

The **Securities and Exchange Commission** is seriously considering building an online tool for bulk downloading of corporate filings. Access to these data would allow analysts to enhance their firm-specific longitudinal records. Again, this information could be integrated into the LBD or NETS.

Business Development and Entrepreneurship

Maryann Feldman, the principal investigator for this project, oversees the maintenance and expansion of the nation's most robust multi-sourced dataset that follows the emergence of technology-intensive

firms in one region over time—the **Bill Little Company Database** covering the Research Triangle, NC area from 1962 through latest year, with real time data collection. The database has detailed longitudinal information on over 3,000 firms, such as start-ups, spin-offs, acquisitions, mergers, business financing, government funding, job change, patent filings, and founder history. Data sources are eclectic and include news media feeds, NETS, corporate reports, government program records, patent data and social media. With the creation of new innovative datasets (such as the SEC tool noted above), opportunities to expand the scope of the database will continue to grow. Most importantly, the reliance on digital data and social media sources means that the data collected are up to the minute. There is a potential using big data methodology to understand complex phenomenon and to also better forecast future trends.

The availability of a Bill Little-type database for economic region would greatly enhance EDA's ability to ascertain the factors that influence the outcomes of EDA-funded assistance to firms. Compared to the high value for economic development research, practice, and evaluation, the cost of funding a set of regional company databases across the nation is relatively small. To test the value of the Bill Little Company Database for evaluation, EDA could consider constructing a pilot

evaluation to test this methodology for other regional economies. EDA could consider a variety of ways in which it might catalyze the creation of such databases, e.g., through information dissemination and a small amount of matching funding, perhaps out of the Regional Innovation Program.

Given the absence of a nationwide set of Bill Little-type regional databases, the next best publicly available resource for tracking individual new firms is **Dow Jones VentureSource**, which provides detailed information on 30,000 venture capital-backed firms in the U.S. It would be interesting to determine if VentureSource (which was not at the data fair) could construct a longitudinal database that includes VC-backed firms that no longer exist. S&P Capital IQ probably could be adapted for this purpose as well—its coverage would need to be compared to that of VentureSource. (**VentureDeal** and **Growthink Research** have much smaller databases of 9,000-10,000 U.S.-based venture capital-backed firms, and so are, at least initially, less attractive resources.)

A number of commercial firms provide access to databases of private equity deals, including venture capital financing and mergers and acquisitions. It would be useful to identify the extent to which EDA investments led to firms developing to the point that

they attracted private equity investors. Resources that can provide data on private equity deals include **CB Insights, S&P Capital IQ, Growthink Research**, and **VentureDeal**; the **MoneyTree Report** (a collaboration between Price Waterhouse Coopers and the National Venture Capital Association) tracks only venture capital deals.

A complementary resource to the prior set is the **Initial Public Offering (IPO) longitudinal database** maintained by Martin Kenney and Donald Patton of the University of California, Davis. As of May 2012, this database included all 2,766 IPOs issued in the U.S. by emerging growth companies between June 1996 and September 2010. The database, compiled primarily from SEC filings, tracks pre-IPO venture capital involvement, IPO proceeds, and post-IPO employment, revenues, firm survival, acquisitions, and bankruptcies, by industry and geography. For evaluation purposes, access to these data would be valuable in combination with those from other, broader resources such as the LBD and NETS. (The Kenney-Patton database also would be a valuable resource for Bill Little-type regional databases.)

Business R&D and Innovation

The single best source of firm-specific data on business R&D and innovation is the National Science Foundation's **Business R&D and**

Innovation Survey (BRDIS), carried out by the Census Bureau. The annual survey of 40,000 firms covers product and process innovations, R&D expenses and capital expenditures, characteristics of R&D, human resources involved in R&D, and intellectual property and technology transfer.¹ The first year of BRDIS coverage is 2008. BRDIS data are integrated into the LBD, so analysis of business R&D and innovation could include non-BRDIS variables.

The limitation on the use of BRDIS is that the data are confidential—only Census employees and outside researchers with special sworn status have access to the data. As noted earlier, a Census staff person indicates that a workable arrangement to serve EDA needs is feasible and so worth exploring.

The **U.S. Patent and Trademark Office** has created a series of research datasets of patents. While patents are acknowledged as an imperfect measure of innovation, their ready availability certainly is worth including. The LBD will be

¹ The Census Bureau says “A mail-out/mail-back sample survey of approximately 40,000 companies with 5 or more employees. Large companies with known R&D > \$3 million from the previous survey cycle are selected each year from the Business Register. The largest 50 companies in terms of payroll within each state are selected each cycle. Smaller companies with more than 5 employees are stratified by industry and payroll size and selected using Probability Proportionate to Size (PPS) sampling.”

incorporating patent data (one of its first non-Census data sources).

Regarding business innovations, the **Thomasnet.com Product News Room** holds promise as a valuable source of up-to-date information on new product introductions for the business-to-business market. At the data fair, Maryann Feldman spoke with Thomas Publishing Co. staff about this idea, which they were quite interested in.

Workforce

Two exciting innovations have the potential to greatly enhance assessment of the impacts of EDA investments on jobs and earnings. The first is the Census Bureau's **Local Employment Dynamics (LED)** longitudinal database, which links establishment files and employee wage records from state unemployment insurance systems. For every establishment every quarter, the LED has the capacity to identify net job change, hires, separations, turnover, tenure, wages, workforce race and ethnicity, worker gender and age, the wages, industry, and location of workers' prior and subsequent employment, and the distance between home and work. It also may be possible for LED to provide imputed estimates of worker educational attainment. Thus, LED could provide rich, unique

information about outcomes quite important to EDA and Congress—jobs, earnings, and job stability.

As with other Census data, LED records are confidential and only Census employees or external researchers with special sworn status can carry out analysis.

The second workforce data innovation of note is “real-time” labor market information (LMI), created through scraping or “spidering” the web for online job ads and analyzing their content to measure the extent and nature of job demand. While real-time LMI tools are most often used for aggregate analysis (e.g., by geography, occupation, industry), they also can be used to track the number and characteristics of job openings by firm. This type of analysis would enable a new dimension of impact assessment of EDA investments. Vendors that provide real-time LMI services include **WANTED Analytics, Burning Glass, Monster.com, CareerBuilder, indeed, and Geographic Solutions.**

Federal Grants and Contracts

USASpending.gov was created by the Office of Management and Budget in 2007 in response to congressional legislation (the Coburn-Obama bill) requiring an online webtool that would allow the public to see the characteristics of any federal grant

or contract and sub-award and to conduct searches by, for instance, performer, funder, and location. The USASpending.gov database is available to generate a list of federal grants and contracts that may have come about as a result of EDA investments and are worthy of further investigation. The primary value of the contracts dataset is identifying the extent to which EDA-funded business assistance may have led to federal work. The grants database is useful in determining the extent to which EDA assistance to governments and nonprofits built their capacity to attract additional federal funds.

Regional Economic Database

The regional economic database will provide a wide, diverse variety of aggregate data for economic geographies such as economic development districts, metropolitan areas, and counties. It is expected that a good part of the database will include traditional federal subnational economic data such as per capita income, the unemployment rate, average wages, income distribution, and academic R&D. That said, a number of innovative data sources have the potential to add substantial value to the database. Categories include economic performance, structure, and dynamics; business development and entrepreneurship; private equity capital; R&D and

innovation; household socioeconomic characteristics; human capital; prices; and dashboard indicators.

Economic Performance, Structure, and Dynamics

A number of innovative data sources offer unique value for looking at the totality of a regional economy in a number of dimensions such as: how well the economy is performing, the distribution of jobs and earnings by industry, the presence of industry and occupational clusters, and dynamic trends in jobs (such as hires and separations) and establishments (openings, expansions, contractions, closings). Having a deep set of regional characteristics will significantly enhance the utility of performance evaluations.

The two most useful innovative data sources are supported by EDA funds—**Innovation in American Regions** and **Stats America** developed by Indiana University and Purdue University. Stats America serves as a comprehensive intermediary of traditional federal data and appears to be an excellent platform on which to build a regional economic database for the purposes of performance evaluation. Innovation in American Regions (IAR) offers a series of useful indicators— an Innovation Index and its 22 components, such as average high-

tech employment share, percent of adult population with a bachelor's degree, and broadband connections per 1,000 households. The site's sponsors are continually testing the explanatory value of various indicators and revising the index accordingly. In addition, the IAR browser provides access to data on knowledge-based occupational clusters, industry clusters, and a multitude of more traditional measures such as educational attainment, per capita income, and labor force size. (It is likely that the traditional measures in Stats America and IAR overlap to some degree.)

Less immediately useful for EDA's purposes are commercial data packagers EMSI and Moody's Analytics. Stats America and IAR carry much the same data, are freely available, were built expressly for the purposes of economic development analysis (using in part data from EMSI), and include unique indicators built from the data.

Harvard Business School's **Cluster Mapping Project**, also EDA-funded, has the potential to be a significant adjunct to IAR. At present, the two projects are in discussion about aligning their efforts.

A potentially useful resource for the economic database is the Brookings Institution's **Export**

Nation, an estimation of the annual export of goods and services for each of the nation's 100 largest metropolitan areas. A decision for inclusion should be made in light of any issues caused by lack of nationwide coverage and the likelihood that Brookings will continue to produce the data series. It would be useful if Brookings could find a way to provide estimates of substate geographies.

Three innovative federal data sources have the potential to provide an overview of job and establishment dynamics for regions. The Census Bureau's **Business Dynamics Statistics** (BDS), drawn from the LBD, has the capacity to construct customized tables by region (1976-latest year) that identify the number of jobs gained through openings and expansions and lost through contractions and closures, by characteristics such as industry and firm age and size. BDS also can count the number of establishments in each of those categories. The Bureau of Labor Statistics' **Business Employment Dynamics** (BED), built on a longitudinal version of the Quarterly Census of Employment and Wages (QCEW), is able to provide a relatively similar type of analysis.

At present, neither the BDS nor the BED provides tables for geographies below the state level, though each organization argues that they plan to do so.

Consequently, the manager of an EDA regional economic database would need to seek customized tables. One problem in the comparability of the two datasets is that there are significant differences in industry classification by establishment, largely because the Census Bureau by law cannot share a good part of its data with BLS. Efforts are underway to rectify this problem.

The Census Bureau's **Local Employment Dynamics** program, discussed earlier, is an excellent resource for generating unique regional measures and indicators. Examples include job churn within an industry; workforce composition by gender, age, and race/ethnicity, by industry; earnings differentials among these various worker categories; the extent of upward economic mobility within the region; the extent of worker migration into and out of the region; and worker access to appropriate jobs within a certain commuting distance. Census has the capacity to construct customized tables for the EDA database.

Business Development and Entrepreneurship

Both **Business Dynamics Statistics** and **Business Employment Dynamics** can be used to generate the rates of new firm formation and survival, with associated jobs, by region.

Narrowing the scope, **DowJones VentureSource** could produce regional data on the number and characteristics of venture capital-backed firms, normalized by regional size. Similarly, CB Insights, S&P Capital IQ, Growthink Research, and VentureDeal could provide regional figures on the number and nature of private equity deals.

Private Equity Capital

Several vendors maintain detailed databases of the providers of private equity by location. **S&P Capital IQ** provides detailed information on over 18,000 private equity firms worldwide (with, presumably, a good percentage in the U.S.). **Private Equity Info**'s database is smaller, 2,000 world-wide, and so not as attractive at first glance.

R&D and Innovation

Data on business R&D and innovation are particularly important to include in the regional economic database. However, at present, the National Science Foundation's **Business R&D and Innovation Survey** has somewhat limited capacity to provide publicly available data by region. NSF is creating experimental estimates of business R&D by metropolitan area. The availability of figures for smaller areas is curtailed to a substantial degree because of Census non-disclosure rules. It would be

worthwhile to more clearly ascertain the extent to which BRDIS can cover U.S. regions.

Elsevier's **SciVal** and Thomson Reuters' **Research Analytics** have similar capacities to access the research strengths of individual universities and laboratories through analyzing publications and citations. (Both firms have extensive worldwide databases of publications and citations, Scopus for Elsevier and Web of Science and Web of Knowledge for Thomson Reuters.) Either data source can be used to prepare regional indicators of academic research strengths, by discipline.

The data fair effort identified a number of efforts to measure the impacts of university R&D, but these are not close to ready for use in a regional economic database. They include the federal interagency **STAR METRICS**, UCLA's **COMETS**, the Association of Public and Land-Grant Universities' **University Economic Impact Metrics**, and the Association of University Technology Managers' proposed **Institutional Economic Engagement Index**.

The U.S. Patent and Trademark Office's **Research Datasets** will allow simple patent metrics by geography. Work within Harvard's **Patent Network Dataverse** provides regional measures of

collaborative and social networks among patent holders.

Household Socioeconomic Characteristics

The Census Bureau's **American Community Survey**, the innovative successor to the traditional decennial long form, can generate annually updated regional measures of a wide range of household socioeconomic characteristics, such as income, occupation, poverty, educational attainment, and reliance on government assistance.

Human Capital

As noted, **Innovation in American Regions** prepares data on occupational clusters around the U.S. If Congress were to fund improvements in the Employment and Training Administration's **Occupational Information Network** (O*NET) to make it more current, IAR could use an occupation-skills crosswalk to map skills clusters as well.

Measures on the strength of a region's postsecondary institutions to prepare its students for well-paying jobs can be generated in the not-too-distant future by states funded under the National Center of Educational Statistics **Statewide Longitudinal Data Systems Grant Program**. The **National Student Clearinghouse Research**

Center says it will have a similar capacity in the future, but details are not available at present.

Prices

The Bureau of Economic Analysis' **Regional Price Parities** (RPP) effort shows the relative cost of living across U.S. regions, overall and for particular classes of goods and services. RPP data, which are about to be produced on a regular schedule, would be a useful addition to a regional economic database. At present, **PriceStats**, which gathers price data by scraping the web world-wide, has less utility for performance evaluation because it does not have the coverage, consistency, and reliability that the RPP will. (Although, the RPP might benefit by adding PriceStats as a secondary data source.)

Dashboard Indicators

The University of Toronto's **Local Indicators for Economic Analysis** (IDEAS) database is a useful resource for identifying indicators for an EDA regional economic dashboard.

Data Tools

In addition to innovative data sources, the data fair also had exhibits on innovative data tools. Open data platforms and visualization and analysis tools are of particular interest for the EDA performance evaluation effort.

Open Data Platforms

Open data platforms provide public web access to datasets from multiple sources. As such, they have the potential to make available previously unseen datasets for incorporation into the third-party microdatabase or the regional economic database.

Data.gov, a federal effort headed by the Office of Management and Budget, is one such platform—to date, it has created “communities” on the topics of business, cities, education, energy, health, manufacturing, and sustainable supply chains.

Private sector platforms with the potential to play the same kind of function include **Windows Azure Marketplace**, **Factual**, **Infochimps Data Marketplace**, and **Amazon Web Services**

Visualization and Analysis Tools

Indiana University’s **Science of Science (Sci2) Tool** is a remarkably useful resource for analyzing and visualizing the contents of all three of EDA’s evaluation databases, including that generated by grantees. ESRI’s **GeoIQ** provides sophisticated data mapping and visualization tools.

Innovative Data Sources

total·Impact

Altmetrics Total-Impact

Categories: R&D, innovation, commercialization; data analysis and visualization tools

Overview: Total-Impact measures the impacts of scholarly research by gathering data on the frequency with which an item of research is tweeted, saved, blogged, downloaded, and bookmarked. In doing so, Total-Impact presents a broader, timelier picture of impact that can complement traditional approaches tracking peer-review or citation analysis. Impact data are aggregated into a single streamlined, shareable report, which can be accessed via an open API and embedded into web-based CVs or article management systems.

Unit of Analysis: Scholarly products, including articles, papers, datasets, slides, and software

Coverage: Scholarly research in 15 Web-based data sources, such as CrossRef, Mendeley, and PLoSALM

Size: Number of unique scholarly products referenced in data sources

Form: User-generated metrics

Key Data Elements: Impact measures vary by data source. Examples include readers, views, citations, mentions, downloads, bookmarks, blogs, tweets, and recommenders.

Data Collection Method: Distributed data system, involving 15 Web-based information sources

Access: Web app is free. Subscriptions are available for high volume usage.

Potential Uses for Economic Analysis: This source compares the research impacts of academic departments and institutions within a region and across regions, by research field or topic. In contrast to more traditional sources, Total-Impact provides an understanding of societal impact.

For Additional Information:

- Website:
 - Total-Impact
 - Altmetrics
- Contacts:
 - Jason Priem (jp@jasonpriem.org)
 - Heather Piwowar (hpiwowar@gmail.com)



Amazon Web Services

Category: Big data, open data platforms, web services

Overview: Amazon Web Services (AWS) offers scalable access to remote IT infrastructure services (cloud computing). AWS offers a number of cloud-based services of use to regional economic analysts and data providers, including dataset hosting and computing, public dataset access, and data search, cleaning, and verification.

Scope: Currently, Public Data Sets on AWS provides access to 13 economic and geographic data sets. Its capacity to host such datasets is close to limitless.

Form: User-determined

Access: Free access to public datasets. Hosting and computing services available on a usage basis.

Potential Uses for Economic Analysis: Data providers can use AWS to host, compute, and analyze big data sets, resulting in cost savings from not having to invest in computer infrastructure. Providers can choose to make these datasets freely available through Public Data Sets on AWS. Researchers also can make use of AWS services to analyze very large data sets. Data users can download existing data from Public Data Sets on AWS. All can obtain data search, cleaning, and verification services through the Amazon Mechanical Turk.

For Additional Information:

- Website: Amazon Web Services
- Contacts:
 - Frank Digiammarino (frankdig@amazon.com)
 - Steven Halliwell (shall@amazon.com)
 - Lue Ray (lueann@amazon.com)
 - Doug VanDyke (Vandyke@amazon.com)



American Community Survey

Category: Jobs, workforce, education and labor markets; Regional industries & economies

Overview: Every month, the U.S. Census Bureau conducts the American Community Survey (ACS) to create detailed data on the U.S. population and how they live. The uses of this database are primarily for the purposes of public policy. ACS summary data, published annually as one-, three-, and five-year averages, are predefined tabulations of socioeconomic characteristics. The basic unit of analysis is a specific geographic entity, ranging in size from block group to the nation, for which estimates of persons, families, households, or housing units in particular categories are provided.

Unit of Analysis: Individuals and households

Coverage: U.S. residents

Form: User-defined tables and analyses

Key Data Elements:

- Age
- Sex
- Race
- Income & benefits
- Family & relationships
- Education
- Health insurance
- Veteran status
- Language spoken at home
- Journey to work
- Occupation
- Rent or ownership status
- Type of housing unit
- Monthly housing costs
- Vehicles available

Geographic Areas: Nation, states, metropolitan areas, counties, places, census tracts, block groups

Timeframe: 2005 to prior calendar year

Frequency: Annual—one-year averages for geographies with 65,000 or more residents, three-year averages for geographies with 20,000 or more residents, and five-year averages for geographies with less than 20,000 residents

Timeliness: 9-12 months after latest reference year

Data Collection Method: 3.5 million addresses sampled each year, with paper form mailed out and non-response follow-up through telephone and in-person interviews

Access: Free, tables through American Fact Finder and datasets through the ACS website

Potential Uses for Economic Analysis: ACS offers useful information on workforce metrics, which include occupation measures and journey-to-work characteristics, with substantial disaggregation by age, gender, race, and ethnicity.

For Additional Information:

- Website: American Community Survey



Business Dynamics Statistics

Categories: Big Data; business creation and development; labor markets; longitudinal databases; regional industries & economies

Overview: Business Dynamics Statistics (BDS) provides measures of net and gross job flows, including measures of job creation and destruction, associated with entering, exiting, expanding, and contracting establishments. Aggregate statistics are available for the nation and states, by firm characteristics and industry classification.

Unit of Analysis: Establishment

Coverage: Matches County Business Patterns coverage

Sectors Covered:

- Agricultural services, forestry, and fishing
- Mining
- Construction
- Manufacturing
- Transportation and public utilities
- Wholesale trade
- Retail trade
- Finance, insurance, and real estate
- Services

Excluded:

- Self-employed
- Domestic service workers
- Railroad employees
- Agricultural production workers
- Most government employees
- Employees on ocean-borne vessels
- Employees in foreign countries

Size: Over 8 million establishment records per year

Form: Longitudinal database, aggregate data tables

Key Data Elements: Establishment openings and closings; firm startups; job creation and destruction by firm size, age, industrial sector

Geographic Areas: Nation, state, and metro/nonmetro (planned)

Industry Detail: Sector (e.g., Construction, Manufacturing)

Timeframe: 1976-2010

Frequency: Annual

Timeliness: Available 16 months after reference year

Methodology: Compiled from the Census Bureau's Longitudinal Business Database, which is constructed by linking annual snapshot files from the Census Bureau's Business Register

Access: Aggregate data tables for public and microdata records for qualified researchers through the network of secure Census Bureau Research Data Centers

Potential Uses for Economic Analysis: With a focus on occupational activity, BDS provides access to data to ascertain patterns of entrepreneurship, structural change, the gross job flows that underlie net employment change, and employment contributions by firm size and age.

For Additional Information:

- Website: Business Dynamic Statistics
- Contact: CES.BDS@census.gov

Business Employment Dynamics

Categories: Big Data; business creation and development; labor markets; longitudinal databases; regional industries & economies

Overview: Business Employment Dynamics (BED) provides quarterly data on establishment openings, closings, expansions, and contractions by industry and size of firm, as well as establishment births, deaths, and survival by age. The data is available at the federal level and by state. BED data are generated from longitudinally linked microdata collected by the Quarterly Census of Employment and Wages (QCEW, formerly the ES-202) program.

Unit of Analysis: Establishment

Coverage:

- QCEW data covers all employers subject to state and federal unemployment insurance (UI) laws—approximately 97% of all U.S. reported employment data
- BED data covers the private sector (excludes government and private households)

Size: 6.8 million establishments; 105 million employees

Form: Longitudinal database, aggregate data tables

Key Data Elements: Gross job gains (expansions, openings); gross job losses (contractions, closings); job losses/gains available by industry sector, firm size, births, and deaths

Geographic Areas: National and state—future expansions may include MSA and county level

Industry Detail: 2- and 3-digit NAICS industry (national), 2-digit NAICS industry (state)

Timeframe: 1992 – Present

Frequency: Quarterly

Timeliness: Available 8 months after reference quarter

Data Collection Method: QCEW collects employment and wage data from quarterly establishment reports submitted to State Workforce Agencies in compliance with unemployment insurance laws. BED assigns a unique identifier to track each business in the longitudinal database.

Access: Multi-screen data search, pre-formatted top picks, FTP site flat file

Potential Uses for Economic Analysis: BED provides data for analysts to identify patterns of gross job creation and destruction by industry sector, and to track survival and identify contributions of young and old business establishments to employment growth.

For Additional Information:

- Website: Business Employment Dynamics
- Email: BDMinfo@bls.gov
- Phone: 202-691-6553



Business R&D and Innovation Survey

Category: R&D, innovation, and commercialization

Overview: The Business R&D and Innovation Survey (BRDIS) provides data on a range of R&D activity performed by U.S. companies by major industry, line of business or business segment, state, and firm size. The National Center for Science and Engineering Statistics (National Science Foundation) and the Census Bureau oversee the administration of the BRDIS. In contrast to the earlier Survey of Industrial Research and Development, BRDIS includes service firms and adds new data elements about innovation.

Units of Analysis: Firms

Coverage: Non-farm, for-profit, public, or private companies, with five or more employees operating in the U.S.

Data are collected for the geographic location of the R&D activity (including foreign locations by country and domestic locations by state and SMSA)

Sample Size: Nationally representative sample of about 40,000 companies, including companies in both manufacturing and nonmanufacturing industries

Form: Annual sets of aggregate statistical tables

Detail: R&D activity by industry, state, and firm size are available for 1953-2007 from the SIRD and for 2008 forward from BRDIS. Data are historically available by state. Experimental data for large metros will be forthcoming.

Timeframe: 2008 – latest year

Key Data Elements:

Financial measures of R&D activity

- Detail on domestic U.S. R&D and on worldwide R&D activity of U.S. R&D performers
- Capital expenditures for R&D

Measures related to R&D management and strategy

- Share of R&D devoted to social sciences, new business areas, and to specific application areas
- R&D partnerships by sector (universities, companies, government) and by type of organization (customer, exhibitor, competitor)

Measures related to R&D employment

- R&D employee headcount by occupational category, sex, and level of educational attainment
- Number of U.S. R&D employees working under visas (H-1B, L-1, etc.)

Measures of company R&D activity funded by organizations not owned by the company

- Worldwide R&D activity and domestic U.S. activity funded by outside organizations
- R&D funded by outside organizations by "business segment" (i.e., below the company level) and by foreign versus domestic organization

Measures related to intellectual property (IP), technology transfer, and innovation

- Participation in activities to introduce new or to improve existing goods, services, methods of manufacturing, distribution, or support systems
- Patent-related data
- Licensing to outside parties
- Participation in specific technology transfer activities

Frequency: Annual

Data Collection Method: BRDIS is structured to encourage different experts within a single business to provide responses in their areas of expertise. Respondents are asked to allocate their domestic and worldwide sales and R&D totals among multiple business codes. Core R&D expenditure questions are intended to provide a bridge between the historical time series and BRDIS. A variety of new questions address data needs identified by users and by businesses themselves.

Access: Public access to tabulations from NSF; restricted access to microdata files from the Census Bureau's Center for Economic Studies

Potential Uses for Economic Analysis: BRDIS allows for the comparison of business R&D and innovation activities by sector and geographic area over time. As BRDIS has elements adapted from the European Union's Community Innovation Survey (CIS), it can be used to make international comparisons on industry R&D and innovation as well.

For Additional Information:

- Website: BRDIS
- Contact: Ray Wolfe (rwolfe@nsf.gov)



Burning Glass: Labor/Insight

Category: Labor markets

Overview: Burning Glass Technologies is a leading provider of labor market analytic and career exploration solutions. A management-owned company founded by scientists, Burning Glass controls several active or pending patents in the fields of data extraction, information interpretation, behavioral profile generation, entity matching, and machine learning.

Burning Glass has applied advanced technologies for collecting and reading free text information from online job ads to create a web-based reporting tool, Labor/Insight. Labor/Insight allows the user to query its comprehensive database of job posting information extracted from over 17,000 online job boards, newspapers, and employer sites on a daily basis. Users can use Labor/Insight to analyze changing employer demand for occupations, skills, education, and certification requirements. Labor/Insight can also be used to identify new and emerging jobs and industries, and changes in individual employer hiring demand within and across sectors.

Labor/Insight differentiates itself from products which utilize keyword text searches or O*Net code searches by applying Statistical Natural Language Processing to mine job posting texts to create an expanded data record that includes skills, education, certification, and salary information in addition to traditionally captured information on occupation, employer, industry, and location.

Units of Analysis: Job posting; regional, state, county and city job markets; specific occupation, job title, skill, or educational credential selected for analysis

Coverage: National, state, city and county job markets, based on more than 17,000 online job-posting sources

Size: Data from approximately 15 to 16 million unique online job listings collected annually

Form: Database of online job postings

Key Data Elements: Data record elements include job function, employer industry, location, education, certification, and skill requirements, and normalized salary. Analysis tools identify geographic-specific job market demands and existing and emerging skill and credential requirements

Timeframe: From 2007 to the present

Frequency: Real-time tracking

Data Collection Method: Patented technology that aggregates, extracts, codes, and normalizes job data from more than 17,000 job boards, newspapers, employers, and other websites

Access: Subscription

Potential Uses for Economic Analysis: Labor/Insight provides (i) profiles of strategic and growth sectors, analysis of existing or lagging skills concentrations, and employer targeting strategies, and (ii) workforce training support and suggestions for alignment to match specific demands for skills in the regional job market.

For Additional Information:

- Website: Burning Glass
- Contact: Michael Cox, Director of Enterprise Sales and Solutions
- Phone: t +1 (617) 227-4800 x 120; m +44 (0) 7870 523024
 - Email: mcox@burning-glass.com



CB Insights

Category: Business Creation and Development

Overview: CB Insights uses big data to track private company financing and M&A data. The company's proprietary machine learning technology tracks venture capital, angel, private equity and government-backed private companies, their investors, and acquirers. With National Science Foundation support, CB Insights is building algorithms that mine public data to assess the health of private companies.

Unit of Analysis: Company & investor profiles

Coverage: Tracks data on high-value private companies ranging from industrial to internet, manufacturing to mobile, and biotech to business services. Covers early-stage companies from those funded by angel investors, government grants, and incubators to private equity and venture-capital-backed firms, as well as under-the-radar, mid-market private companies. Investors include venture capital and private equity firms, state and federal grant programs, individual angel investors, and angel groups as well as incubators and accelerator programs.

Size: Data on over 80,000 high-value private companies and 24,000 investors and acquirers

Form: Micro-level database of financing deals, M&A and IPOs

Key Data Elements:

Firm Level Detail

- Industry
- Geography
- Name
- Keyword
- Funding History
- Management Team
- Competitors
- Funding Events
 - Venture Capital
 - Angel investments
 - Private Equity

Timeframe: Tracks venture capital deals back to 1999 and angel and government financings back to 2007. Private company M&A activity goes back to 2007.

Frequency: Data is updated on a real-time basis every day.

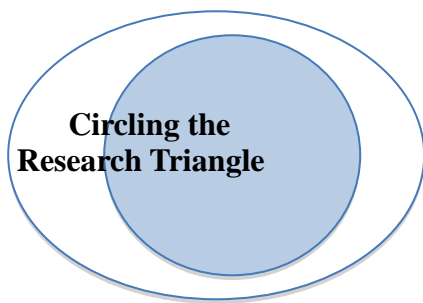
Data Collection Method: CB Insights aggregates data via machine learning technology it has developed that parses structured entity information from unstructured, semi-structured, and structured information sources. On a daily basis, the company's technology crawls SEC filings, news publications, social media, investor and company websites and tens of thousands of other sources identifying investment and M&A news related to private companies of interest.

Access: Subscription required for current data; historical VC and angel reports are available for free

Potential Uses for Economic Analysis: Economic development groups ranging from NYCEDC to the Government of Singapore use CB Insights to achieve a few objectives, namely to (i) identify sectors/industries of growth to inform their economic development agendas and plans; (ii) benchmark their regions against other areas; (iii) target companies who have raised money to establish a presence in their region; (iv) identify investors and acquirers for local companies looking for growth capital. In addition, this source aids local companies in search of investment. This ultimately creates more jobs and increases local tax base among other benefits.

For Additional Information:

- Website: CB Insights
- Contact: Anand Sanwal (asanwal@cbinsights.com)



Circling the Research Triangle

Categories: Business creation & development; longitudinal database; regional industries & economies

Overview: This database matches a variety of data sources useful for studying the industrial genesis of the region surrounding the Research Triangle in North Carolina. The objective is to develop a robust platform for integrating diverse data sources to provide insights into regional industrial development. Maryann Feldman and Nichola Lowe, along with UNC students, have created a relational database that follows individual firms over time and also provides educational background data and career histories on founders.

Unit of Analysis: Establishment

Coverage: Universe of entrepreneurial starts-up and establishments in technology-intensive industries (e.g., life sciences, information and communication technology, gaming, cleantech and business services) in the 13-county North Carolina Research Triangle Park region from 1962 to the present

Size: More than 4,100 establishments

Form: Database

Geographic Details: 13-county Research Triangle region as designated by the Research Triangle Regional Partnership

Key Data Elements:

- Year of incorporation for startups
- Year of relocation for established firms formed outside the region
- Sector, subsector & technology
- Complete address
- Corporate affiliations, if applicable
- Annual Employment
- Annual Sales
- Annual Patent filings
- Participation in business development programs and initiatives
- Key financial milestones, such as
 - Venture capital infusion
 - Federal small business assistance financing
 - State grants & awards
- Liquidity events, such as
 - IPO
 - Acquisitions
 - Mergers
 - Bankruptcy
- Educational attainment and career history of founders (for startups)

Timeframe: Annual from 1962 to the present

Frequency: Collected continuously & still under development

Data Collection Method: Original data collection & synthesis

Access: Currently available upon request

Potential Uses for Economic Analysis: The database allows for an in-depth understanding of the complexity of the process of regional economic change and the role of constituent organizations over time. The underlying data collection methodology and database structure may be replicated in other regions.

For Additional Information:

- Website: Circling the Triangle
- Contact: Maryann Feldman (maryann.feldman@unc.edu)



Cost of Living Adjustments for the New Supplemental Poverty Measures

Category: Prices and costs

Overview: The Interagency Technical Working Group (ITWG) on *Developing a Supplemental Poverty Measure* recommended that Supplemental Poverty Measure thresholds be adjusted for price differences across geographic areas. The American Community Survey (ACS) data allows for the first time the creation of a housing price index. Only the shelter portion of the Supplemental Poverty Measurement thresholds is adjusted using the index. This approach offers one option for calculating cost-of-living differences across regions.

Unit of Analysis: Two-bedroom rental units with complete plumbing and kitchens

Coverage: For each state, a median is estimated for all nonmetropolitan areas (48), for each metropolitan statistical area (MSA) with a population above 100,000 (264), and for a combination of all other metropolitan areas. There are currently 358 index values.

Key Data Elements: The BLS, using data from five years of the Consumer Expenditure survey, estimates thresholds for renters, homeowners with mortgages, and homeowners without mortgages. The index is based on median gross rent for two-bedroom units with complete plumbing and kitchens. Additional estimated indices consider rent at the 33rd percentile and below poverty thresholds.

Form: Index

Timeframe: Currently 2009

Frequency: Annual

Data Source: Derived from the 5-Year American Community Survey data

Access: Public use, available on the SPM Research Files

Potential Uses for Economic Analysis: For an understanding of long-term economic impacts, this data source provides for cost-of-living comparisons across regions or within states with consideration to metropolitan/non-metropolitan areas.

For Additional Information:

- Website:
 - Available on the SPM Research Files
 - Development of Index: Geographic Adjustments of Supplemental Poverty Measure Thresholds: Using the American Community Survey Five-Year Data on Housing Costs
- Contact: Trudi Renwick, U.S. Census Bureau
301-763-5133
Email: trudi.j.renwick@census.gov



Data.gov

Category: Big data, open data platforms, web services

Overview: The purpose of Data.gov is to increase public access to high value, machine-readable datasets generated by the Executive Branch of the Federal government. As a priority, the Open Government Initiative for President Obama's administration, Data.gov increases the ability of the public to easily find, download, and use datasets that are generated and held by the Federal government.

Data.gov provides descriptions of the Federal datasets (metadata), information about how to access the datasets, and tools that leverage government datasets, enabling users to directly analyze the underlying information. Data.gov is committed to creating an unprecedented level of openness in government. This openness aims to strengthen our Nation's democracy and promote efficiency and effectiveness in Government.

Coverage: Data aggregated from 172 Federal, Executive Branch agencies and sub-agencies are included in Data.gov

Scope: 390,834 raw and geospatial datasets

Form: Open data platform

Key Data Elements:

- *Raw Data Catalog:* a catalog with instant view/download of platform-independent, machine-readable data. Links to a metadata page have additional links to authoritative source information from the sponsoring agency's website including pertinent agency technical documentation regarding the dataset.

- **Geodata Catalog:** trusted, authoritative, Federal geospatial information. This catalog includes links to download the datasets and a metadata page with details on the datasets, as well as links to more detailed Federal Geographic Data Committee (FGDC) metadata information.
- **Tools Catalog:** simple, application-driven access. This catalog features widgets, data mining and extraction tools, applications, and other services.

Timeframe: The data catalogs will continue to grow as datasets are added.

Frequency: Updated continuously

Access: Public; multi-screen data search, pre-formatted top picks, or FTP site flat file

Potential Uses for Economic Analysis: Agencies have been asked to post datasets on Data.gov that increase government accountability by revealing the results and characteristics of government services to citizens; the public's use of government services; the distribution of funds from the government; and demonstrable results from Federal programs. All of these are crucial elements of accountability. The efficiency of information-centric markets benefits economic development directly by ensuring maximum access to available information. The Open Data Set movement has spread to 31 U.S. States and 15 American Cities. There are 30 international sites with similar access, enabling greater comparability.

For Additional Information:

- Website: Data.gov
- Email
- Phone: 800-333-4636



Dataverse Network Project

Category: R&D, innovation, and commercialization

Overview: Created and hosted by the Institute for Quantitative Social Science at Harvard University, the Dataverse Network Project is a virtual web archive that allows researchers to publish, share, reference, extract, and analyze research data. This is a flexible platform to allow researchers to manage their data while maintaining credit and ownership, managing updates and granting access to others. Lee Fleming has used this tool to organize and share data on matched patents and publications.

Form: This open source application provides a personalized data archive platform that allows users to upload, manage, and protect their data.

Access: The application is free to the public. Access to the data in the dataverse is determined between the manager of the dataset and the party interested in gaining access.

Potential Uses for Economic Analysis: This research tool provides an open source application for publishing, citing, and discovering research data related to a number of R&D areas that allow users to share data relevant for regional economic and innovative analysis.

For Additional Information:

- Website: [Dataverse](https://dataverse.harvard.edu/)

- Contact: Vetle Torvik (vtorvik@illinois.edu)



DataWeb for the DataFerrett

Category: Data intermediaries and integrators

Overview: The U.S. Census Bureau offers DataFerrett as a data mining, extraction, and analytical tool, allowing users to locate and retrieve data, select and recode variables, and develop and customize tables, graphs, and maps to create visual depictions of the data. DataFerrett sources data from TheDataWeb, a distributed network of public and private databases providing a vast amount of statistical information that is constantly updated and expanded. The DataWeb team collaborates with cross-agency data providers, as well as public sector partners to enhance and extend the DataFerrett project.

Form: Data platform. Available types of datasets include microdata, aggregate or summarized data, longitudinal datasets, and time series datasets.

Scope: Ninety-four supplements and modules, and thousands of individual monthly, quarterly, and annual releases; 97 search topics range from Adult School Enrollment to County Population Estimates

The following datasets are available, and within those, all supplements or modules and most releases:

- American Community Survey
- American Housing Survey
- Behavioral Risk Factor Surveillance System
- Consumer Expenditure Survey
- County Business Patterns
- Current Population Survey
- Decennial Census of Population and Housing
- Home Mortgage Disclosure Act
- National Ambulatory Medical Care Survey
- National Center for Health Statistics
- National Health and Nutrition Examination Survey
- National Health Interview Survey
- National Hospital Ambulatory Medical Care Survey
- National Survey of FHWAR
- NYC Housing and Vacancy Survey
- Public Libraries Survey
- Small Area Health Insurance Estimates
- Small Area Income and Poverty Estimates
- Social Security Administration
- Survey of Income and Program Participation
- Survey of Program Dynamics

Frequency: Release dates occur monthly, quarterly, annually, or periodically, depending on the dataset.

Method: Multiple surveys and their subsequent supplements and releases are made available through TheDataWeb on an on-going basis. Datasets are linked to TheDataWeb network, and accessed via the DataFerrett tool.

Access: Public

Potential Uses for Economic Analysis: TheDataWeb is a useful tool for analysts to enhance accessibility and usability of relevant datasets and integrate of data from multiple sources.

For Additional Information:

- Website: TheDataWeb for the DataFerrett
- Contact:
 - Bill Hazard (William.g.hazard@census.gov)
 - Rebecca Blash (Rebecca.v.blash@census.gov)



Discovery Logic

Categories: Networks and relationships; R&D, innovation, and commercialization

Overview: Discovery Logic *connects the dots* across scientific and business databases to inform R&D investment decisions, visualize trends, locate experts and evaluate project and portfolio performance. Specializing in mining and refining knowledge from large scientific databases, Thomson Reuters connects research to impact. Subject matter experts deliver custom content, metrics and indicators, tools and systems, and interpretive studies and reports that support producers, funders and publishers of research and scientific information.

The global platform, *ScienceWire®*, aggregates content and related people, products, organizations and outcomes. In addition, it combines search and data mining technology, advanced algorithms and inter-source linkages among real-time open-source and proprietary databases including cross-agency R&D grant data, patents, citations, journals and news. They maintain and update these databases, create and apply algorithms to extract and link information across the databases, and deliver custom solutions that allow clients to apply the data to address their business needs.

Units of analysis: includes experts, research products, and research organizations

Coverage: Contents of source databases, including:

- Publication and topic data from *Web of Science*SM and NLM MEDLINE
- Grant award data from the NIH, NSF, DOE, DOD, USDA, and NASA
- Patent data from the US Patent and Trademark Office
- Derwent Patent Data

Form: Analytic tool

Access: Subscription

Potential Uses for Economic Analysis: Discovery Logic is a tool that enables users to create scientific decision support systems. More specifically, this tool provides a platform that allows analysts to utilize advanced algorithms and inter-source linkages among real-time open-source and proprietary databases including cross-agency R&D grant data, patents, citations, journals and news.

For Additional Information:

- Website: Discovery Logic—Thomas Reuters
- Contacts:
 - Matt Probus (matt.probus@thomsonreuters.com)
 - Elizabeth Deitz (elizabeth.deitz@thomsonreuters.com)



EMSI Analyst

Categories: Data analysis & visualization tools; labor markets; regional industries and economies

Overview: Economic Modeling Specialists Incorporated (EMSI) provides web-based software and consulting services built around labor market data. The data they develop comes from over 80 state and federal sources. EMSI pulls it together, cleans it up, and presents it so clients can use it. The database contains comprehensive information on industries, occupations, and demographics — as well as data on occupational skills, education, training, and specific regions and industries.

Units of Analysis: Jobs, workers, residents, establishments

Coverage: Data provided by the U.S. Department of Commerce, U.S. Department of Labor, Employment and Training Administration, U.S. Department of Education, National Center for Education Statistics, U.S. Department of Health and Human Services, National Center for Health Statistics, U.S. Postal Service, Internal Revenue Service, U.S. Railroad Retirement Board, and private sources that include Indeed.com and Nielsen Claritas Business-Facts.

EMSI augments federal data by filling in suppressions and including the self-employed, agricultural workers, and others not captured by basic payroll data.

Geographic Detail: Nation, U.S. county, MSA or zip code

Form: Aggregate labor market data tables based on approximately 80 sources

Key Data Elements:

Industry Data

- 2-6 Digit NAICS Industries
- 2001-Present Historic Data
- 10-year Projections
- Regionalized Staffing Patterns by Industry
- Average Annual Earnings
- Establishments
- Unemployment by 2-Digit NAICS
- Occupation Data

Full Detail SOC Occupations

- 2001-Present Historic Data
- 10-year Projections
- Regionalized Staffing Patterns by Occupation
- Average Annual Earnings
- Median and Percentile Earnings
- Unemployment by 2-Digit SOC

Demographics

- Population by Age, Race/Ethnicity, and Gender
- Educational Attainment (current and projected)

Other

- INDEED – job postings linked to occupations
- Businesses (by name and size) tied to NAICS codes
- Career Clusters
- O*NET (Occupational Competencies)
- IPEDS
- Patents

Timeframe: 2000-present as well as 10-year projections

Frequency: Updated quarterly

Access: Subscription

Potential Uses for Economic Analysis: The data are used to assess and understand employment trends, education and economic development, and dynamics.

For Additional Information:

- Website: EMSI
- Contact: Rob Sentz (rob@economicmodeling.com)



Export Nation 2012

Category: Regional industries and economies

Overview: The Brookings Metropolitan Policy Program's Export Nation 2012 provides a large database of geographically-detailed international export data, goods and services, which are estimated by location of production. While the U.S. Census Bureau produces a state exports series and prepares a metropolitan export series for the International Trade Administration, these series reflect origin-of-movement export data limited only to merchandise exports. The origin-of-movement, however, is not always the place where the good was produced, especially when the exported goods are consolidated along the shipment route.

Units of Analysis: Aggregate exports

Key Data Elements: Exports, by export destination and major industry, including:

- Nominal and real exports, total and by industry (major and detailed)
- Exports share of Gross Domestic Product
- Direct export-production jobs, total and by major industry
- Total export-supported jobs, total and by major industry
- Annualized real export growth rates, total and by industry (major and detailed)

Geographic Detail: The export data is available for each of the 3,113 counties in the U.S.; 366 metropolitan statistical areas (metros); 576 micropolitan statistical areas (micros); 50 states plus the District of Columbia; and the United States.

Industry Detail: There are 34 major industrial categories analyzed: 26 for goods exports (3-digit level NAICS) and eight

categories for services (U.S. Bureau of Economic Analysis service export categories). In addition, the dataset provides export data for 212 subcomponents of the major export industries, both goods and services.

Form: Report and web-accessible tables.

Method: The appendix of the “Export Nation 2012” study provides a detailed explanation of the methodology for constructing this data series.

Timeframe: 2003 to 2010

Frequency: Annual

Data Sources: United States International Trade Commission, the Bureau of Economic Analysis, the Bureau of Labor Statistics, the Internal Revenue Service, Moody’s Analytics, and NAESA: The Association of International Educators.

Access: The detailed data are available as a series of aggregated data tables at this link. A series of indicators for the largest 100 metropolitan areas for 2010 can be accessed with this interactive tool on the Brookings Metro Program website.

Potential Uses for Economic Analysis: To understand economic activity within a broader, global context, this data source enables analysts to identify the role of international exports in local economies, unveil the export industrial base at the local level, and determine a metropolitan area’s role in the global marketplace.

For Additional Information:

- Website: Brookings Metropolitan Policy Program
- Contacts:
 - Emilia Istrate, Associate Fellow (eistrate@brookings.edu)
 - Nicholas Marchio, Research Assistant (nmarchio@brookings.edu)



Factual

Category: Big data, open data platforms, and web services

Overview: Factual is an open data platform for application developers that leverage large-scale data aggregation and community exchange. The focus is on making data more accessible (i.e. cheaper, higher quality, and less encumbered) for machines and developers to drive and accelerate innovation.

Unit of Analysis: Varies depending on data source

Scope: Factual aggregates data from many sources including partners, user community, and the web.

Global Places

- Global Places Database
- U.S. Healthcare Providers Extended Attributes
- U.S. Restaurants Extended Attributes
- Place Crosswalk—map of places across the web

Global Product

- Global Products Database
- Products Crosswalk—map of products across the web

Value Added: Factual applies a sophisticated machine-learning technology stack to extract both unstructured and structured data; clean, standardize, canonicalize data; and merge, de-dupe, and map entities across multiple sources.

Form: Open data platform

Timeframe:

- Real-time data
- Time-series materializations may be possible depending on data set and use case

Frequency: Updated in real-time

Potential Uses for Economic Analysis: By making data more accessible, Factual provides access to clean, structured data that enables analysts to customize analyses of multiple facets of regional economic activity using web sourced data.

For Additional Information:

- Website: Factual
- Contacts:
 - Vikas Gupta (vikas@factual.com)
 - Leo Polovets (leo@factual.com)



GeoIQ

Categories: Big Data, open data platforms, web services; data analysis and visualization tools

Overview: GeoIQ is a client-based geospatial data management, visualization, and analysis platform. Users can share and merge data, while using location as the common pivot point to identify trends and patterns, fuse together large amounts of information from numerous data sources, and identify trends and opportunities to drive better business decisions. Data and maps are shared through GeoCommons, a public platform to which GeoIQ users contribute location-relevant information.

Form: Geospatial analysis tool

Units of analysis: User-determined

Coverage: GeoIQ relies on user-submitted data on which geospatial data analysis is performed.

Key Functions: Allows for the use of many geospatial analytical techniques, including aggregation, prediction within and across data sets, filtering, option for custom equations creation, temporal analysis of time-based data

Products

- GeoIQ Explorer
- GeoIQ Geocoder
- GeoIQ Acetate
- GeoIQ Mobile
- GeoIQ Social
- Pro Services
- Dev Tools

Size: More than 27,000 active users have mapped over 508,000

data layers and used more than 2.5 million data sets to create 9.5 million maps

Frequency: More than 27,000 users regularly upload data sets

Access: Data published to be publicly available is analyzed for free; proprietary data analysis is conducted on a subscription-basis

Potential Uses for Economic Analysis: Geospatial data analysis can track a diverse set of location-specific economic and social factors, including infrastructure, population concentration, demographics, employment density, and other regional development indicators.

For Additional Information:

- Website:
 - GeoIQ
 - GeoCommons 2.0

Green Goods and Services Survey

Categories: Labor markets; regional industries and economies

Overview: The Green Goods and Services (GGS) survey provides a measure of national and state employment in industries that produce goods or provide services that benefit the environment. The GGS is included within the Quarterly Census of Employment and Wages (QCEW).

Form: Pre-defined tables

Unit of Analysis: Establishment

Coverage: The GGS survey includes business and government establishments within 333 industries that are identified as potentially producing green goods or providing green services. The sampling frame, the QCEW, covers 98 percent of U.S. jobs available at the county, MSA, state and national levels by industry. GGS fall into one or more of the following groups:

- Production of energy from renewable sources
- Energy efficiency
- Pollution and greenhouse reduction or recycling and reuse
- Natural resources conservation
- Environmental compliance, education and training, and public awareness

Size: 120,000 establishments

Key Data Elements: Green jobs, industry shares of green jobs

Geographic Detail: State level employment estimates by 2 digit NAICS are available.

Timeframe: Began in calendar year 2010

Frequency: Annual news release with descriptive tables and quarterly web-only updates

Data Source and Collection Method:

- The BLS QCEW provides GGS with establishment employment data. Self-employed workers are not included in the BLS count of green jobs.
- A company's share of revenue from green products is used to estimate a company's "green employment."

Access: Free

Potential Uses for Economic Analysis: In an effort to qualify economic activity at the state level by focusing on environmentally friendly economic activity, this data source provides useful information on the scope of green goods and services.

For Additional Information:

- Website: GGS: Green Goods and Services
- Contact: Richard Clayton, 202-691-6515, clayton.rick@bls.gov



Innovation in American Regions

Categories: Data intermediaries and integrators; regional industries and economies; R&D, innovation, and commercialization

Overview: Much of today's successful economic growth hinges on attracting or cultivating jobs that characterize the "innovation economy"—firms and occupations relying on talented workers whose skills are based on knowledge, insight, and creativity. Innovation-based economic growth in rural America, however, has long lagged behind the nation's metropolitan areas. To address this gap, the U.S. Economic Development Administration sponsored this project to develop new tools to support strategic economic development planning in rural regions.

The goal of this work is to help rural planners and development practitioners assess their regions' comparative strengths and weaknesses with respect to fostering innovation-based growth. While the primary focus of the project was to help rural regional planning, the project's data and tools are equally well-suited for any type of geographic definition—urban, exurban, metropolitan or user-defined—depending upon the practitioner's specific need and purpose.

Form: Interactive data tools

Geographic Detail: County and user-defined multi-county regions.

Key Data Elements:

- Educational attainment, population characteristics, establishments, employment & wages, housing & households, income, earnings & poverty, labor force (LAUS)
- County-level data on 15 knowledge-based occupation clusters and 17 industry clusters are also available in this interactive database. Analytical tools help regional planners evaluate public investment decisions in support of economic growth
- An Innovation Index reflecting a region's innovation activity and capacity, together with an interactive database containing the index and its component indicators for every county in the nation

Data Sources:

- | | |
|---|---|
| • Economic Modeling Specialists Inc. (EMSI) | • Internal Revenue Service |
| • Federal Communications Commission | • The Innovation Index uses data from the above government statistical agencies and private, proprietary sources. |
| • Innovation Economy 360, Decision Data Resources | • The industry clusters are built with QCEW and IBRC estimates for undisclosed values |
| • Moody's economy.com | • The occupation clusters are provided by EMSI and Purdue University |
| • National Science Foundation | |
| • U.S. Bureau of Economic Analysis | |
| • U.S. Bureau of Labor Statistics | |
| • U.S. Census Bureau | |

Timeframe: Data span from 2001 to the present, depending on the data series

Frequency: Data are updated when the sources release new data. The Innovation Index is updated periodically.

Access: Free to the public

Potential Uses for Economic Analysis: The mapping tool allows users to easily compare innovation capacity and industry and occupation clusters in different counties and regions—both “official” and user-defined geographic definitions—around the nation. This helps in assessing the relative strengths and weaknesses of the region’s clusters. The drill-down feature for the Innovation Index allows the users to view and download the non-proprietary data used to calculate the index and its components.

For Additional Information:

- Website: Innovation in American Regions
- Contacts:
 - Carol Rogers (rogersc@iupui.edu)
 - Timothy Slaper (tslaper@indiana.edu)

Kenny-Patton IPO Database

Kenny-Patton IPO Database

Category: Business creation and development

Overview: This database is comprised of all *de novo* initial public offerings (IPOs) on American stock exchanges and filed with the Securities and Exchange Commission (SEC) from June 1996 through December 2006.

Derived from Thomson Financial Venture, the following types of firms and filings were excluded: mutual funds, real estate investment trusts (REITs), asset acquisition or blank check companies, foreign F-1 filers, all small business (SB-2) IPOs (to be added in Version 2.0 released in August 2012) with the exception of Internet firms, and all spin-offs and other firms that were not true *de novo* firms (such as, firms formed purely to acquire other firms, etc.).

Unit of Analysis: Corporations

Coverage: Firms with *de novo* IPOs on American stock exchanges and filed with the Securities and Exchange Commission (SEC)

Size: 2,500 firms and over 25,000 individuals

Data Source: SEC's Electronic Data, Gathering and Retrieval (EDGAR) website

Form: Database

Timeframe: June 1996 through December 2006. Version 2.0 released in August 2012 provides data for 2006-2010.

Access: Free upon email request (mfkenney@ucdavis.edu)

Key Data Elements:

Firms

- Name
- Locations (street address, city, state, zip code)
- Exchange and ticker
- Auditor
- Year of founding & year of incorporation

Firms Managers/ Firm Directors

- Basic individual data
- Previous positions and previous firm (incomplete due to difficulty of clearly identifying these in all cases)
- Year of joining the firm
- Education data

Lawyers

- Name
- Addresses of law firms

Underwriters

- Name of lawyer
- Name and address of law firms representing the lead investment banker

Offering

- Date of IPO
- Share Volume
- Initial Share Price
- Shares Outstanding at time of IPO
- Underwriter discount

New data elements being added:

- 2006-2010 and SB-2 filers
- Firm employment by year for every year after the IPO through 2010
- Firm revenues by year for every year after the IPO through 2010
- Firm outcome after IPO through 2010 (i.e., continuing, merged or acquired, or bankruptcy)

Potential Uses for Economic Analysis: In an effort to understand firm activity beyond firm formation, this database provides information on IPO activity. More specifically, this database allows analysts to examine spatial patterns of IPO activity and proximity of external agents that support firms undertaking IPOs.

For Additional Information:

- Website: IPO Database
- Contact: Martin Kenney (mfkenney@ucdavis.edu)



Local IDEAs (Indicator Database for Economic Analysis)

Category: Regional industries and economies

Overview: Local IDEAs (Indicator Database for Economic Analysis) is constructed to support a wide range of statistical analysis as a central resource for a network of researchers. The database includes an extensive set of social and economic indicators that contribute to the economic vitality of Canadian localities to benchmark their performance against other countries, particularly the U.S. The indicators included in the database are in the process of being assembled through a combination of publicly available sources. In addition, special tabulations of economic and social data are available through the purchase of complementary private sets of data. Local IDEAs is a project of PROGRIS: Program on Globalization and Regional Innovation Systems at the Monk School of Global Affairs, University of Toronto.

Units of Analysis: City-region, defined as the presence of a core city linked by functional ties to a surrounding hinterland based on travel to work

Form: Aggregate data tables

Timeframe:

- Municipalities: standardized from 1986 forward
- Regions: some indicators standardized from 1971 forward, most from 1996 forward
- Business records: 2001; 2006; 2011; Patents: 1975-2010

Frequency: Annual for income, patents, R&D, business patterns; years ending in 6 or 1 for the Census of Population and D&B full file businesses

Access: Free

Key Data Elements:

- Demographics
- Labor market
 - Occupation levels
 - Educational attainment levels
 - Academic fields of study
 - Unemployment
- Immigration and domestic migration
- Canadian business pattern data
- Canadian business records (approximately 1.5 million records)
 - 8-digit SIC/ 4-digit NAICS
 - Employees
 - Revenues
 - Full contact details
- Global corporate ownership structure
- Profiles of industrial clusters (19 types by 144 regions + flexibility to create custom types)
- City/region GDP estimates
- Detailed geo-reference data on Canadian patent filers
- Local private R&D expenditure data from Impact Group
- Detailed geo-references
- Public R&D funding

Data Sources:

- Statistics Canada
- Dun & Bradstreet
- Hoover's
- USPTO
- Canadian Association of University Business Officers (CAUBO)
- Conference Board of Canada
- Innovation Atlas
- Canada Revenue Agency
- University Spin-offs

Potential Uses for Economic Analysis: To understand the economic potential of a region, it is critical that we understand the greater social and economic environment that supports this activity. PROGRIS has data on the social and economic factors that provide the foundation of prosperity for city-regions.

For Additional Information:

- Website: PROGRIS
- Contact: David Wolfe (david.wolfe@utoronoto.ca)



Longitudinal Business Database

Category: Business creation and development; longitudinal databases

Overview: The Longitudinal Business Database (LBD) is a research dataset constructed at the Center for Economic Studies (CES) in the U.S. Census Bureau. The LBD contains the universe of all U.S. business establishments with paid employees listed in the Census Bureau's business register. Updated annually, the LBD provides data on all employer establishments that are in scope for the Economic Census, as well as a large number of other out-of-scope entities. The LBD provides researchers with a complete and accurate set of longitudinal establishment linkages, and contains basic information on establishment size, payroll, age, industry, location, ownership, and legal form of organization as well as characteristics of the firms they belong to. This includes firm age and firm size. The LBD can be linked to other establishment and firm information contained in Economic Census and survey files available at CES.

Form: Microdatabase, panel series

Unit of Analysis: Establishment

Coverage: All U.S. business establishments with paid employees listed in the Census Bureau's business register

Key Data Elements: Establishment size, industry, location, ownership, start year, last year

Size: 8 million observations in 2009 & 24 million unique establishments from 1975 to present

Timeframe: LBD: 1976-2009

Frequency: Annual

Method: Constructed from linkages between establishments across annual Standard Statistical Establishment List files

Access: Use restricted to qualified researchers, through the Research Data Center Program

Potential Uses for Economic Analysis: To understand trends of firm formation, tenure and death, LBD provides data to track the evolution of region's establishments by location or industry. Additionally, this data source allows for comparative analysis of the contributions of young vs. old firms or small vs. large firms.

For Additional Information:

- Website: LBD
- Contact: Email ces.contacts@census.gov



Local Employment Dynamics

Category: Big data, open data platforms, and web services; labor markets; longitudinal databases; regional industries and economies

Overview: The Longitudinal Employer-Household Dynamics (LEHD) database consists of linked employer-employee data from 49 states and the District of Columbia. The underlying data is a complex system of linked state unemployment insurance wage record data and Quarterly Census of Employment and Wages (QCEW) data, linked to Census demographic and business data. The state data is made available to Census from the Local Employment Dynamics (LED) state-federal data-sharing partnership.

As most statistical products are calculated either from a household frame or a business frame, a linked employer-employee jobs frame has enormous potential to provide new information about the economy.

Currently two public use data products are derived from the LEHD data, the Quarterly Workforce Indicators (QWI) and OnTheMap. The QWI are 30 labor force indicators, providing detailed information on employment, job creation/destruction, and wage dynamics by worker demographic characteristics (age, sex, education, and race/ethnicity). The OnTheMap synthetic data allows for the mapping and reporting of employment and home locations of workers within user-defined areas. The LEHD program at Census continues to work on new public use data products developed from the LEHD jobs data, and to enhance the existing set of data products.

Nature of Source: Microdata

Unit of Analysis: Job

Coverage:

- UI-covered employment only; federal and self-employed workers to be added
- All states except Massachusetts and Puerto Rico and the Virgin Islands

Size: The LEHD microdata are extremely large, covering all UI-covered jobs for 49 states over the available time series

Key Data Elements: QWI labor force indicators include total employment, net job flows, job creation/job destruction, new hires/recalls, separations, turnover, and average monthly earnings (all workers, new hires, attached workers)

Geographic Detail:

- QWI is released at the state, county, metro, and WIA level
- OnTheMap/LODES is released at the Block level

Timeframe: Time series availability varies by state; several states have data back to the early 1990s, some states not available until mid-2000s

Frequency: The underlying microdata are quarterly data. QWI is updated quarterly, and OnTheMap (calculated off of Q2 data) is updated annually.

Data Collection Method: Collects no additional data; state partners supply UI wage & QCEW records & WIA geographic definitions; records are linked with Census demographic and business data

Access: QWI and OnTheMap are public use data products. LEHD microdata are confidential but can be accessed by researchers with approved projects through the secure Census Research Data Centers (RDCs).

Potential Uses for Economic Analysis: QWI allows for comparative analysis of workforce metrics by providing data on demographics and wages of newly hired versus other workers in the same industry and employment trends at sub-state geographies. In addition, OnTheMap can identify residential concentrations of workers in the local labor market.

For Additional Information:

- Website: Longitudinal Employer-Household Dynamics
- Contact: Erika McEntarfer, Lead Economist, LEHD Research, Center for Economic Studies
Erika.McEntarfer@census.gov



Mendeley

Categories: Networks and relationships; R&D, innovation, commercialization

Overview: Mendeley is a free reference manager and academic social network that helps users organize their research, collaborate with others online, and discover the latest research. The desktop component is a workflow tool used to organize, read, annotate, and cite papers, individually or in collaboration with colleagues. The web component includes a public database of research. Readership statistics are aggregated and tracked, providing real-time data on the usage of papers within the Mendeley network. Users with shared interests may join public groups. This can further enhance collaboration opportunities by highlighting popular or new works or identifying potential research collaborators.

Form: Analytical tool

Unit of Analysis: Research references

Coverage: More than 60 million unique papers covering all academic disciplines, ranging from Arts and Literature to Mathematics, Physics and Computer Sciences; approximately 1.8 million users worldwide

Key Data Elements: Paper downloads and user profile visits

Size:

- 1.8 million users
- 170 million papers uploaded
- 40,000 public groups

Timeframe: Includes research references from the past 100 years

Frequency: Continuous

Data Collection Method: Users import papers into their own personal library; Mendeley extracts the meta-data and adds these records to its online database of research. Mendeley tracks readership of each paper and displays aggregated results

Access: Free for individuals; premium packages available for teams and individuals; Institutional Edition, powered by SWETS

Potential Uses for Economic Analysis: Mendeley may be used as both a data source to track regional research activity and as a collaborative platform to enhance regional collaboration, which can result in added economic value.

For Additional Information:

- Website: Mendeley
- Contact: community@mendeley.com

MONSTER REAL-TIME LABOR INTELLIGENCE

ACTIONABLE MARKET INFORMATION, ANALYSIS AND INSIGHTS

Monster Real-Time Labor Intelligence

Category: Labor markets

Overview: Monster's real-time labor intelligence (RLI) offers timely data, analysis and insights to drive key program and investment decisions. Government, education and business customers use the research services and analysts to make decisions that grow industries, create new employment opportunities for job seekers, and help align workforce skills with employer needs.

Talent Dashboard uses a fundamentally new approach to understanding resumes with Monster's 6Sense semantic search technology. Unlike the other products, it understands concepts and context to provide an unrivaled level of understanding and accuracy. 6Sense patented search uses a combination of complex search algorithms coupled with a comprehensive knowledge base to understand the concepts and context in a resume.

Unit of Analysis: Segment by job title, experience, skill, location, education, school, company, degree, and age of resume

Coverage: Over 100 million U.S. resumes and 650K new resumes added monthly. Monster manages over 22 million active Job Seeker Accounts

Key Data Elements: Semantically parsed from Monster's resume postings or private resume databases, not dependent on structured user input

Form: Micro database

Timeframe: Current resumes from past 12-24 months

Frequency: Continually updated

Data Collection Method: Data is compiled by users of Monster.com or from resume databases loaded into the platform for analysis

Access: Hosted cloud solution

Potential Uses for Economic Analysis: Until now, talent pools were simply collections of resumes. Monster provides in-depth, detailed labor market data to provide the critical insight for business attraction, expansion, and retention; site selection; talent identification and attraction; talent supply analysis and precision talent matching; developing workforce strategy, and policy.

For Additional Information:

- Websites
 - Monster
 - Real-time Labor Intelligence
- Contact for RLI: Bruce Stephen
(Bruce.Stephen@monster.com)
- Contact for SeeMore: Javid Muhammedali
(Javid.Muhammedali@Monster.com)

Moody's Analytics

Category: Data intermediaries and integrators; prices and costs; regional industries and economies

Overview: Moody's Analytics is a leading independent provider of data, analysis, modeling, and forecasts on national, state, metro, and county economies. Their staff of 65 economists and 25 data specialists collects time series of historical economic data and create forecasts for key indicators down to the detailed regional level.

Form: Aggregate data tables

Units of Analysis: Includes firms, establishments, workers, jobs

Scope: Data from public sources and from partner organizations such as Equifax, LPS, Corelogic, NAR, and Case Shiller among others.

Geographic Detail: National, state, metro, counties

Key Data Elements: Topics includes banking/financial, consumer credit, demographics, price/interest rates, industry/labor/employment, housing/real estate

Functionality: Includes data download, charting and mapping, and automating report generation

Timeframe: History goes back as far as the source goes and forecasts are for 30-year projections

Frequency: Historical data is updated within hours of being released from the source. Baseline and five alternative scenario forecasts are updated monthly.

Access: Historical and forecast data are available on a subscription basis through DataBuffet, a web-based interface that allows downloads to most software formats. Access to the company's staff of 65 economists is included with every subscription.

Potential Uses for Economic Analysis: While most research focuses on retrospective analysis, this data source allows analysts to not only look at historical trends, but also to analyze multiple dimensions of regional economies and forecasting.

For Additional Information:

- Website: Moody Analytics
- Contact: Robin Heid
 - Email: Robin.Heid@moody.com
 - Phone: (610) 235 5186.



National Establishment Time-Series (NETS)

Categories: Longitudinal database; regional industries and economies

Overview: Walls & Associates converts Dun and Bradstreet (D&B) archival establishment data into a time-series database of establishment information, the National Establishment Time-Series (NETS) Database. The NETS Database provides longitudinal data on dynamics of the U.S. economy.

Unit of Analysis: Establishment

Coverage: Business, non-profit and government establishments, sole proprietors

Size and Timeframe: 44.2 million unique business, non-profit and government establishments between 1990 and 2010

Form: Micro-level database

Key Data Elements:

- Business establishments
- Headquarter linkages
- Relocation information
- National count of related establishments
- Years active
- Industry classification
- Employment
- Estimated annual sales (firm-level)
- Dun & Bradstreet Credit score
- Special indicators
- Dun & Bradstreet PayDex scores

- Type of establishment

Frequency: Annual update

Method: Annual “snapshots” of D&B’s proprietary establishment data are utilized to construct the NETS Database time-series. No establishments are deleted from the database; the “First Year” and “Last Year” are provided to indicate which are still active in 2010.

Access: Subscription

Potential Uses for Economic Analysis: The data provides access to analysis on the following topics:

- What is the size and performance of specific markets over time and do we want to invest in them?
- How has a specific firm’s (or set of firms) market share changed over time?
- Who are the important employers in a region and who contributes most to the region’s growth?
- What kinds of occupations are going to be in demand?
- Business startup and failure analysis.
- What kinds of linkages are there among businesses in your state?
- What are the impacts of tax changes, environmental regulations, and educational performance on business location decisions?
- Do establishments that receive venture capital perform better than those that do not?
- Product line forecasting.
- Epidemiology studies of the links between industries and disease.
- Comparisons of large and small retail chains and their competition with independent stores.
- Job creation and destruction at the industry level by establishment.
- Does local ownership foster growth?
- Economic development targeting.
- How does public policy impact business performance?
- Do business incubators foster business success and survival?

For Additional Information:

- Website: NETS
- Contact: Don Walls (dwalls2@earthlink.net)



National Student Clearinghouse

Categories: Labor markets; longitudinal databases

Overview: The National Student Clearinghouse Research Center collaborates with higher education institutions, states, school districts, high schools, and educational organizations to better inform education leaders and policymakers. The Research Center focuses on longitudinal data outcomes reporting. The *StudentTracker* and *StudentTracker for High Schools* tools allow for institutions of higher education or high school districts to track student performance, enrollment, and graduation data.

Unit of Analysis: Student

Coverage and Size: 3,300 colleges and universities, enrolling over 94% of all students in public and private U.S. institutions

Form: Longitudinal microdatabase

Timeframe: Enrollment records are generally available from 2000 onward. Degree coverage is available from 1990 onward.

Frequency: Enrollment and degree information is obtained every 30-45 days

Data Collection Method: Participating collegiate institutions provide enrollment and degree records. StudentTracker matches records across the institutions' submissions to provide annual updates on current, former, or prospective students.

Key Data Education Outcomes That Can Be Tracked:

StudentTracker College and University

- Real time reporting on all currently enrolled students
- Ability to track transfer and student persistence in postsecondary education
- Students graduating each year
- Students moving from an undergraduate degree into graduate programs
- Interstate college student mobility

StudentTracker for High Schools:

- Allows high schools, school districts, regional consortia and states to follow the enrollment activities of graduates, including:
 - Immediate or delayed college/university enrollment
 - Persistence, degree attainment, & time to college graduation
 - Potential for comparative benchmarking reports on performance of high schools and school districts

Access: No anonymized data sets are available for open research use. Access to student level data is limited to directory information. Researchers must contact the NSC Research Center and provide information on the purpose, scope, and feasibility of the research to be granted access.

Potential Uses for Economic Analysis: NSC Research Center provides access to education data, which offers an important lens for understanding the strength and potential of a workforce in a given region or industry. This data includes information on university and college enrollment. Time to graduate and persistence metrics can help capture the in- or out-migration of high school graduates, as well as the concentration of college graduates within a region.

For Additional Information:

- Website: National Student Clearinghouse Research Center
- Email: researchdirects@studentclearinghouse.org



O*NET Data Collection Program

Category: Labor markets

Overview: O*NET, the Occupational Information Network, is a comprehensive database of worker attributes and job characteristics. O*NET supports public and private sector efforts to identify and develop the skills of the American workforce. It provides a common language for defining and describing occupations. Its flexible design reflects the rapidly changing job requirements.

Unit of Analysis: Occupation

Coverage and Size: The O*NET Data Collection Program provides information for over 900 O*NET-SOC occupations, covering the entire U.S. economy. These occupations can be directly linked to the Standard Occupational Classification (SOC) system.

Key Data Elements: The database contains information about knowledge, skills, abilities (KSAs), interests, general work activities (GWAs), and work context. Each occupation has 239 descriptors and over 400 ratings. O*NET can link related occupational, educational, and labor market information databases to the system.

Timeframe: Ongoing since 2002

Frequency: The project has established a continuing data collection program to populate and maintain the O*NET database. Approximately 100 occupations are updated annually.

Data Collection Method: The O*NET questions have been organized into several different questionnaires covering various aspects of the occupation. A minimum of 60 surveys are completed for each occupation. Data are collected primarily from sampled workers. A subset of occupations has ratings collected from occupational experts. Workers sampled from establishments are randomly assigned to answer only one of three questionnaires. Occupational experts complete all three surveys. In addition, trained occupational analysts provide skills and abilities information. Information collection thus far has included over 40,000 businesses and 170,000 employees.

Access: Available for download (tab-delimited text format; SQL format; Microsoft Access; SAS/PC versions) and directly to end users via the following websites: O*NET Online, My Next Move, My Next Move for Veterans, and O*NET Code Connector.

Potential Uses for Economic Analysis: The skill-base and workforce are critical for economic activity. O*NET data allows analysts to assess regional occupational and skills clusters, using crosswalk between occupations and skills; identify workforce KSAs and GWAs; and explore evolution of regional job base over time. See O*NET Resource Center for examples of O*NET at work.

For Additional Information:

- Website: O*NET
- Contact:
 - David Rivkin (rivkin.david@dol.gov)
 - Phil Lewis (lewis.phil@dol.gov)



PriceStats

Category: Big data, open data platforms, web services; prices and costs

Overview: PriceStats is a leading source of inflation statistics, using online prices to develop daily inflation figures across multiple economic sectors in 70 countries.

Alberto Cavallo and Rigoberto Rigobon at MIT founded PriceStats in September of 2010 as a spin-off of the Billion Prices Project, an academic initiative that leverages online prices to conduct academic research related to inflation and price behavior. PriceStats currently brings its inflation series to the financial sector through a strategic partnership with State Street Global Markets. State Street distributes PriceStats' daily inflation updates to their clients through their proprietary website, IR3.

Nature of Data Source: Micro-level data

Unit of Analysis: Prices

Coverage and Size: 5 million products sold by 700 retailers in 70 countries. Product categories including food and beverage, clothing, housing, recreation, household products, and health.

Key Data Elements: Includes price, product description, product attributes, sale indicator and out of stock indicator. Country inflation series contain daily averages of individual price changes across multiple categories and retailers, by sector

Timeframe: 2007 to present

Frequency: Daily

Data Collection Method: PriceStats uses a variety of software to collect price data from online retailers and then uses advanced econometric models to create inflation indices.

Access: Free 10-day lagged US Index. All other indices require subscription via State Street Global Market's proprietary website IR3. PriceStats also works with organizations on an ad-hoc basis to develop customized tools and statistics that can improve decision making related to public policy or pricing strategies.

Potential Uses for Economic Analysis: This data allows analysts to monitor price trends and provide real-time information on inflation. This offers a tangible metric for assessing economic activity in a region.

For Additional Information:

- Website: PriceStats
- Contact:
 - Email: contact@pricestats.com
 - Phone: (617) 577-3908



Research Data Centers

Categories: Business creation and development; regional industries and economies

Overview: The Center for Economic Studies (CES), part of the U.S. Census Bureau, provides restricted access to longitudinal data from the Business Register, which is compiled from the quinquennial Economic Census and business surveys.

Units of Analysis: Firm and establishment

Coverage: Varies on survey and data analyzed

Size: Over 8 million business records in 2009

Form: Longitudinal microdatabase

Timeframe: Varies on survey and data analyzed, many predate 1980

Frequency: Depends on dataset—updated monthly, quarterly, annually, or every five years

Additional Datasets Available for Linkage: Census demographic and decennial data. Patent and export datasets expected.

Access: Researchers must apply for access to the Research Data Centers (RDC) to gain special sworn status from the Census.

Potential Uses for Economic Analysis: RDC provide a rich source of micro-level data on economic activity within the US. With proper access, analysts can examine research topics that

include but are not limited to the impact of trade, venture capital financing, and rural entrepreneurship.

For Additional Information:

- Website: RDC Research Opportunities
- Contact: Local RDC administrator where data access will occur

Regional Price Parities

Category: Prices and costs

Overview: Regional Price Parities (RPPs) are price indexes that measure the price level differences between places for one time period.

Unit of Analysis: Expenditure Class

Coverage: Food, apparel, recreation, transportation, housing, education, medical, other goods and services as well as rents

Geographic Detail: State, metro, county

Timeframe: 2005-2009

Form: Pre-defined tables

Frequency: The release of 2006-11 RPPs is planned for summer 2012.

Access: Excel data tables

Key Data Element: Spatial price index (US = 100)

Potential Uses for Economic Analysis: Regional Price Parities provide data for the comparison of price levels across different geographic areas in general and for specific groups of goods and services. Moreover, analysts have the flexibility to adjust regional measures of income and output to examine price level differences across different industries and communities.

For Additional Information:

- “Regional Price Parities by Expenditure Class for 2005-2009” in the May 2011 issue of the Survey of Current Business
- Program Email: rpp@bea.gov
- Contacts:
 - Eric Figueroa (eric.figueroa@bea.gov)
Phone: (202) 606-9328
 - Troy Martin (troy.martin@bea.gov)
Phone: (202) 606-9207

Rural Establishment Innovation Survey

Categories: R&D, innovation, and commercialization; regional industries and economies

Overview: The proposed survey of rural establishments will primarily assess the adoption of innovative practices and their contribution to firm productivity, the availability and use of local and regional assets, and the extent and importance of participation in Federal, State, and local programs are designed to promote rural business vitality and growth. The survey is immediately concerned with producing an inventory of rural innovation and a comparison with urban establishment. The data may also be linked with Business Employment Dynamics (BED) data to examine associations of innovative behavior with establishment survival and employment growth.

Unit of Analysis: Establishments

Coverage: Metropolitan and nonmetropolitan business with 5 or more employees, active in the following tradable sectors: mining, manufacturing, wholesale trade, transportation, warehousing, information, finance and insurance, professional/ scientific/ technical services, arts, management of business

Sample Size: 30,000 respondents

Form: Aggregate tables

Geographic Detail: Type of geography (e.g., urban, rural)

Key Data Elements:

- Employees and employee backgrounds and training
- Factors affecting location decision
- Utilization of various government and government-sponsored programs
- Financing strategies
- Business challenges related to location
- Technology utilization
- Innovation activity
- Interaction with other businesses by location and type

Timeframe: Data collection is anticipated to end in early 2013

Frequency: Current plans are for a one-time collection

Data Collection Method: Survey of a stratified random selection of establishments. Screening interview will be used to determine the most knowledgeable person in the establishment to respond to the survey. Multi-modal survey instrument (phone, mail and web) and token incentives will be used to increase response rate.

Access: To be determined

Potential Uses for Economic Analysis: Data from the Rural Establishment Innovation Survey provides information to examine associations of innovative behavior with establishment survival and employment growth. In addition, this resource offers information to analyze factors central to economic activity including the role of location (urban versus rural), firm size, local institutions, and proximity in innovation activities and innovation networks.

For Additional Information:

- Website: Rural Establishment Innovation Survey



S&E Indicators State Data Tool

Category: R&D, innovation, and commercialization

Overview: The Science and Engineering Indicators State Data Tool presents 58 indicators that can be explored through tables, charts, and maps. Indicators can be examined in depth or compared to one another. Tables can show up to 20 years of data by year or state with time trends. Charts show indicators over years for all states and can be sorted by year or state. The quartile map shows the geographic distribution of an indicator by year. Comparisons can be made across states or years with varying selection. Indicators include a variety of variables of elementary and secondary education performance, higher education achievement, workforce measures, R&D input and output statistics, and science and technology in the economy measures.

Coverage: Data is available on all states over the last 20 years for 58 indicators of science and technology in education, the workforce, R&D, and the economy.

Form: Aggregate data tables, charts, maps

Key Data Elements:

Elementary & Secondary Education

- Math & science performance & proficiency for fourth & eight graders
- Rates of AP testing
- Public school teacher salaries
- School expenditures as share of GDP & per pupil

Higher Education

- Rates of bachelor's degrees conferred
- Science & engineering degrees ratio
- State funding of universities & student aid

Science & Technology in the Economy

- High tech establishment rates

Financial R&D Inputs

- R&D as percentage of GDP

- High tech employment rates
- Venture Capital investments
- Federal R&D obligations
- State agency R&D expenditures

R&D Outputs

- Science & engineering doctorates
- Academic science & engineering article output
- Academic patents awarded

Workforce

- Science & engineering occupations in the workforce
- Percentages of science & technology careers in the workforce

Geographic Detail: State

Timeframe: Varies by indicator but some date back to 1993

Frequency: Data are updated in accordance with the biennial *Science and Engineering Indicators*

Data Sources: Data sources are listed for each indicator and include NCSES, BLS, Census, NCES, and BEA. Some data in the tool are derived from administrative records from patent offices, and publication records.

Access: Publicly available

Potential Uses for Economic Analysis: Research has shown that S&T activity is critical for innovation, economic activity and growth. NCSES provides access to data for analysts for assess trends, inputs, and outputs in science and technology over time and across states.

For Additional Information:

- Website: S&E Indicators NCSES Website
- Contact: Jeri Mulrow, jmulrow@nsf.gov

S&P **CAPITAL IQ**

S&P Capital IQ

Categories: Business creation and development; networks and relationships

Overview: S&P Capital IQ offers detailed information on public and private capital markets along with applications for desktop research, screening, real-time market data, backtesting, portfolio management, financial modeling, quantitative analysis, and more.

Units of Analysis: Firms, executives, and industries

Coverage: Active and inactive public and private companies, worldwide

Size: Over 88,000 active and inactive companies worldwide, including over 41,100 active public companies

Form: Database

Data Sources: Combines offerings previously provided by Capital IQ, elements of S&P including Global Credit Portal and MarketScope Advisor, enterprise solutions such as S&P Securities Evaluations and Compustat, research offerings including Leveraged Commentary & Data, Global Markets Intelligence, and company and fund research.

Frequency: Updated continuously

Access: Subscription

Key Data Elements:

- Breadth--data elements represent 5,000 unique financial data items, with over 2,500 industry-specific items
- Qualitative data
 - Company Intelligence (including Business Relationships- Subsidiaries, Strategic Alliance Partners, Customers and Competitors)
 - People Intelligence (including Compensation, Stock Ownership and Insider Trading for Public companies, Who knows whom)
 - Key Developments
 - Private Equity
 - Transactions
- Global market data
 - Macroeconomic indicators including employment, GDP, balance of payments, & inflation
- Sell-side Research and Estimates
- Fixed Income
- Alpha & Risk Models
 - Stock Selection Models
 - Equity Risk Models
- Financials and valuation
 - Pricing and Market Data
 - Compustat Financials
 - Equity Capital Structure
- Industry Profile

Potential Uses for Economic Analysis: CapIQ provides a wealth of information on corporate activity, including data on detailed financials and linkages and activities among companies, investors, and subsidiaries in specific regions. This data is not only useful for analysis at the firm-level, but also in understanding individual- and firm-level collaborations

For Additional Information:

- Website: S&P Capital IQ
- Address: 55 Water Street, 49th Floor, New York, NY 10041
 - Phone: +1 212 438 8701
 - Email: information@capitaliq.com

Sci² Tool

Science of Science (Sci²) Tool

Categories: Data analysis and visualization tools

Overview: The Science of Science (Sci²) Tool is a modular toolset specifically designed for the study of science. It supports the temporal, geospatial, topical, and network analysis and visualization of scholarly datasets at the micro (individual), meso (local), and macro (global) levels.

Units of Analysis:

- Micro (Individual)
- Meso (Local)
- Macro (Global)

Type of analysis:

- Statistical/profiling
- Temporal
- Geospatial
- Topical
- Network Analysis

Functionality: Users of the tool can access datasets online or load their own data; perform different types of analysis with the most effective algorithms available; use different visualizations to interactively explore and understand specific datasets; and share datasets and algorithms across scientific boundaries.

The Sci² Tool was designed with scientometric data in mind, but can easily be used to visualize data from other sources. The tool is capable of reading and working with a number of file types, including ISI, GraphML and generic comma-delimited data. It is a

“plug-and-play macroscope” that is compatible with the OSGi industry standard.

Frequency: Periodic—there are two to three new releases of Sci² each year.

Implementation Method: Utilizes the OSGi.org industry standard and the Cyberinfrastructure Shell for integration of algorithms and tools. The Sci² tool team welcomes developers to contribute their own extensions to the Sci² project; see algorithm developer guide.

Access: Open source, Apache 2.0 license

Potential Uses for Economic Analysis: Sci² tool allows for the visualizations of any form of data on individuals and places. With a focus on scientometric data, researchers have the capacity to examine temporal, geospatial, topical, network analysis of scholarly datasets at multiple levels.

For Additional Information:

- Website: Sci²
- Contact: Robert Light (lightr@indiana.edu)



SciENCv: Science Experts Network and CV

Categories: Networks and relationships; R&D, innovation, commercialization

Overview: SciENCv, or Science Experts Network and CV, is a voluntary data platform that will allow the scientific community to document their research activity and maintain pertinent and current CV information. The SciENCv project is closely connected to the [STAR METRICS](#) program.

Unit of Analysis: R&D researchers

Coverage: Researchers who chose to participate in the program

Form: Database, report-generating tools

Key Data Elements: Measures include research expertise, employment, education, and professional accomplishments. Data can be categorized by individual researcher, specific project, and institution.

Method: The data collection in SciENCv will be facilitated by automated feeds from existing data repositories. Information will be claimed and controlled by the users. The system will allow researchers to prepopulate data collections associated with extramural grants and other federally supported research projects. A researcher's SciENCv will describe an individual's scientific contributions and it will allow for discovery of potential partnerships through the open source database. The researcher owns all of their profile data and can control its visibility.

Timeframe: Pilot system to be introduced in early fall 2012.

Frequency: As part of the Star Metrics pilot project, this platform operates as an open source database that is continuously updated.

Access: Open access, any researcher may register in the system. The individual researchers control the visibility of their data.

Potential Uses for Economic Analysis: SciENCV is a rich data source that provides information on expertise, education, networks, and research by an area or institution. This offers valuable information on the workforce skills and human capital.

For Additional Information:

- Website: SciENCV
- Contact: Walter Schaffer (SchaffeW@od.nih.gov)



SciVal, Elsevier

Categories: Networks and relationships; R&D, innovation and commercialization

Overview: Elsevier's SciVal suite of online tools and analytical services provides information about expertise availability, researcher productivity, and funding agency activities and characteristics.

Form: Database

Data Source: Publication data from the Scopus database includes 19,500+ peer-reviewed journals from 5,000+ publishers worldwide and 46+ million records; includes data on 14,000+ active funding opportunities and 2.46 million+ awarded from 4,500+ funding agencies.

Key Data Elements:

Research Institutions:

- Research networking
- Researcher performance
- Funding opportunities and award history
- Bibliometrics, including publication details, citations & downloads

Funding Agencies:

- Funding portfolios of similar agencies
- Funded project spending and performance
- Reviewer identification
- Subject-matter expertise identification

SciVal can also integrate content provided by institutions such as HR data, grants, and patents.

Timeframe: The Scopus database includes 25 million records with references back to 1996 (of which 78% include references) and 21 million records pre-1996 which go back as far as 1823.

Frequency: Constantly updated

Access: Subscription

Potential Uses for Economic Analysis: Assess the value of social capital among a network of individuals or within an institution by examining the strengths, capabilities and research topics of researchers and research institutions. SciVal allows analysts to identify researchers with relevant expertise and interests, examine research awards and explore funding opportunities, and identify research collaborations.

For Additional Information:

- Website: Elsevier SciVal
- Contact:
 - Phone: 1 888 615 4500
 - Email: usinfo@elsevier.com



STAR METRICS

A Federal Collaboration with Research Institutions

STAR METRICS

Categories: Business creation and development; R&D, innovation, commercialization; regional industries and economies

Overview: STAR METRICS (Science and Technology for America's Reinvestment: Measuring the Effects of Research on Innovation) is a partnership between science agencies and research institutions to document science investments and their outcomes to the public. By harmonizing data reporting based on existing systems of record, STAR METRICS provides detailed, systematic information about science investments and their outcomes.

Applications of the STAR METRICS data platform include the Portfolio Explorer Project, a tool to examine public research award information by topic, region, institution, and researcher.

Units of Analysis: University, researcher, research project, federal science agency

Coverage: Federal agencies participating in STAR METRICS include NIH, NSF, DOE, EPA, USDA and the Office of Science and Technology Policy, along with more than 80 U.S. colleges and universities.

Tools: STAR METRICS currently operates four tools to view scientific portfolios:

- The Portfolio Viewer provides information on proposals, awards, researchers, and institutions by program level and scientific topic.

- The Expertise Locator provides information on proposals and co-PIs related to different topic areas to find researchers working on that topic.
- The Patent Viewer provides data on patents from NSF grantees.
- The Map Viewer offers a geographic tool to view NSF investment by institution and topic.

Additional applications of the STAR METRICS platform are under development.

Form: Database, web tool

Key Data Elements: Data on awards (grant topics, funding), employees (occupational code, employment status, earnings from awards), indirect costs, exhibitors (payments from awards), sub-award (payments grouped by sub-award recipient)

Multiple levels of analysis are feasible, such as individual and establishment level data on award recipients, award level data for Federal science agencies, and networks of R&D activity in the public and private sector.

Geographic Detail: For Portfolio Explorer, nation, state, congressional district, institution

Data Collection Method: STAR METRICS combines Federal research award data with de-identified information about individuals, exhibitors, and sub-awards associated with awards. Standardized reporting of core data elements comes from multiple institutions.

Timeframe: Data back to 2008 or earlier for some institutions; other institutions from earliest date of STAR METRICS participation.

Frequency: Internal data updated quarterly; public data updated frequently beginning in 2010.

Access: Limited for database, public for Portfolio Explorer

Potential Uses for Economic Analysis: STAR METRICS enables estimation of the multiplier effects of federal R&D spending. This data is useful for estimating the direct impacts of R&D employment and spending on local exhibitors, and indirect effects through education, employment, research, and network effects.

For Additional Information:

- Website: STAR METRICS
- Contact:
 - John King (USDA) (johnking@ers.usda.gov)
 - Kei Koizumi (OSTP) (Kei_Koizumi@ostp.eop.gov)

Statewide Longitudinal Data Systems

Category: Labor markets; longitudinal databases

Overview: The National Center for Education Statistics (NCES) manages a competitive grants program to support the development of Statewide Longitudinal Data Systems (SLDS) that are intended to enhance the ability of state governments to efficiently and accurately manage, analyze, and use education data based on individual student records. SLDSs help state education agencies, districts, schools, workforce development organizations, and educators and trainers to make data-informed decisions that improve student learning and outcomes and facilitate research to increase student achievement and close achievement gaps. Forty-two states and the District of Columbia have received SLDS grants.

Nature of Source: Micro database

Unit of Analysis: Students

Coverage: Pre-kindergarten through postsecondary education and into the workforce (P-20W)

Key Data Elements: Working with key stakeholders, NCES is developing voluntary standards and guidelines (Common Education Data Standards) to assist state educational agencies in developing SLDSs. Standard data definitions will help ensure that data shared across institutions are consistent and comparable. This, in turn, will make it easier for states to learn how students fare as they move across institutions, state lines, and school levels and how they fare in the workforce.

Timeframe: Varies by state

Frequency: Updated regularly, varies by state

Data Collection Method: Local school districts and postsecondary institutions provide student records to the state. Workforce outcomes of education gathered primarily through use of employee wage records in the state unemployment insurance system. State workforce agencies also can provide records of participants in government-sponsored workforce development programs.

Access: Varies by state

Potential Uses for Economic Analysis: SLDSs allows analysts to trace the workforce outcomes (in terms of employment, industry, and wages) of students' education. Possible comparisons include those by geographic area, demographic characteristics (gender, age, race, and ethnicity), educational institutions, nature of credentials, educational programs, type of curriculum, and veteran status.

For Additional Information:

- Website: Statewide Longitudinal Data Systems Grant Program
- Website: Data Quality Campaign



STATS America

Categories: Data intermediaries and integrators; regional industries and economies

Overview: STATS America is a service of the Indiana Business Research Center at Indiana University's Kelley School of Business that accesses data items from hundreds of data sets from federal and state sources, along with some commercial or private source data. To ensure total accuracy, all data are verified first in analyzing the source of data. STATS America adds value to these data through easy access and functionality, while acknowledging the agency source of the data on every table, profile or map. The database provides access to calculations, graphs, comparisons of time or geography, time series and maps.

Units of Analysis: States, counties & metropolitan areas; flexible Radius Region Builder tool

Form: Aggregate data tables

Key Data Topics:

- Economy
- Education
- Income & taxes
- Population & Housing
- Workforce

Timeframe: Variable, depends on data source

Frequency: Data are updated as they are released from the sources; an online tracks release

Access: Free

Potential Uses for Economic Analysis: STATS data offers information to assess the strengths and weaknesses of regional economies, particularly with regard to brainpower and innovation. This is a useful tool for economic development analysis given that it offers a simple way to calculate whether a county, region, or tract meets certain federal grant thresholds based on unemployment and per capita income.

For Additional Information:

- Website: STATS America
- Contacts:
 - Carol Rogers (rogersc@iupui.edu)
 - Timothy Slaper (tslaper@indiana.edu)



STATT: Statistics Access for Tech Transfer

Category: Business creation and development; R&D, innovation, commercialization

Overview: Statistics Access for Tech Transfer (STATT) is a searchable, exportable database of 20 years of academic licensing data from participating academic institutions collected by the Association of University Technology Managers (AUTM).

Unit of Analysis: Universities, research institution, and teaching hospitals

Coverage: 350+ member institutions of universities, research institutions and teaching hospitals in U.S. and Canada

Form: User-generated spreadsheets from database

Key Data Elements: Disclosures, patent applications filed, patents received, licensing activity and income, startups, commercial products, funding, staff size, legal fees, and more.

Timeframe: 1991 – 2010

Frequency: Annual

Data Collection Method: Annual survey of AUTM member organizations.

Access: Subscription

Potential Uses for Economic Analysis: The university plays a critical role supporting the innovative infrastructure of a region.

The AUTM data provides university-level data that allows for the assessment of institutional technology transfer capacity and contributions by region.

For Additional Information:

- Website: STATT
- Contact: Richard Kordal, Ph.D., RTTP
Director, Office of Intellectual Property &
Commercialization
Louisiana Tech University
+1-318-257-2484
email: rkordal@latech.edu



TEN: The Evidence Network

Categories: R&D, innovation, and commercialization; regional industries and economies

Overview: The Evidence Network (TEN) uses a novel methodology to assess the impact of investments in research, innovation, and business support. The impact assessment architecture allows the team to produce standardized impact assessments that are comparable across organizations and programs and over time. The surveys can be customized to reflect the specifics of program delivery mechanisms and targeted client firms. TEN's standardized yet customizable approach addresses the challenge of benchmarking dissimilar programs.

TEN measures the impact of investments in regional institutions such as economic development organizations, research institutes, university technology transfer offices, research and technology commercialization networks, science parks, and business support programs.

Unit of Analysis: Firms on whose behalf the investments in research, innovation, and business support have been made

Form: Institution and program-specific reports

Key Data Elements:

Resources & Capabilities

- Knowledge, information, advice
- Financing

Firm Performance

- Revenues
- Exports

- Opportunities for promotion & influence
- Research & business linkages
- Technology services
- Complementary business services
- Employment
- Valuation
- Ability to raise financing
- New products
- Process & services
- Time to market

Data Collection Method: Primary, firm level data collected directly from the firms on whose behalf the investments in research, innovation, and business support have been made. Their clients provide contact information that is used for the web-based survey.

Access: Fee-based access

Potential Uses for Economic Analysis: TEN offers data on the measurement of impacts of investments in research, innovation, and business support; data is available by funder, class of funder, and region. Moreover, by standardizing these metrics, cross-regional and cross-organizational comparisons are possible. Results inform future investments and support continuous improvement in the design and delivery of program services.

For Additional Information:

- Website: TEN: The Evidence Network
- Contact: Dr. Brian Barge
President & CEO
The Evidence Network
barga@theevidencenetwork.com



ThomasNet's Product News Room

Category: R&D, innovation, and commercialization

Overview: Thomas is an information and technology company that connects buyers and sellers, primarily business-to-business in manufacturing. ThomasNet is an online platform that combines semantic product search technology with company profiles. The ThomasNet Product News Room generates notice of and information on a very large number of new product introductions in the U.S.

Units of Analysis: Product introductions, companies

Coverage and Size: Database of more than 607,000 companies in the industrial marketplace, categorized by 67,000 product and service classifications

Key Data Elements: New product announcements contain information on product, company, location, and market. Substantial information available on individual firms as well.

Timeframe: Announcements are made available in real time

Data Collection Method: Self-reported by suppliers and verified by ThomasNet editors

Access: Individual announcements free at ThomasNet.com & ThomasNet News. Customized databases available on request, for fee. ThomasNet is open to discussions about providing data in all formats.

Potential Uses for Economic Analysis: Surprisingly, much of the scholarship on economic outputs focuses on job creation, firm establishments, ROI, IP activity, and even publications. This data source offers valuable information on product introductions, which provides a tangible output resulting from firm production.

For Additional Information:

- Website:
 - Thomas Net Website
 - Thomas Net News Website
- Contacts:
 - Paul Gerbino, Publisher, ThomasNet News (pgerbino@thomasnet.com)
 - Linda Rigano, Executive Director, Strategic Services (lrigano@thomasnet.com)

U.S. Cluster Mapping BETA

U.S. Cluster Mapping Project

Category: Data Intermediaries and Integrators

Overview: The U.S. Cluster Mapping Project is an effort of the Institute for Strategy and Competitiveness at Harvard Business School, funded by the U.S. Economic Development Administration. The project aims to provide policymakers and development practitioners with data and tools to assess regional cluster strengths, business environment characteristics, and innovation assets; with case studies on and toolkits for formulating development strategies; and with a directory profiling active cluster initiatives throughout the country. The project's tools are in beta format, under development.

Unit of Analysis: Cluster

Coverage and Size: 41 clusters currently, with plans to map the entire U.S.

Form: Interactive data tools based on industrial clusters, generating aggregate data tables and maps

Key Data Elements:

- Each region's clusters: specialization, employment, wages, job creation, patents
- Comparisons of clusters across the U.S.
- Overall regional economic performance (performance indicators, patents, jobs, wages)
- Characteristics of cluster initiatives

Timeframe: Annual economic data from 1998 to 2010

Frequency: Updated when new underlying data becomes available. Core industrial data for clusters typically becomes available each year in June or July.

Method: Clusters are defined by creating a grouping of standard industry codes, using employment linkages across geographies. Underlying data source is the Census Bureau's County Business Patterns.

Access: Free

Potential Uses for Economic Analysis: The cluster mapping effort provides a unique lens for analysts to examine industrial activity and agglomeration trends within a regional context. This resource offers detailed data on a region's economic structure and allows for systematic comparison across regions. Cluster category definitions are standardized facilitating cross regional comparisons

For Additional Information:

- Website: U.S. Cluster Mapping
- Contacts:
 - Rich Bryden (rbryden@hbs.edu)
 - Samantha Zyontz (szyontz@hbs.edu)



USAspending.GOV

Categories: R&D, innovation, and commercialization; regional industries and economies

Overview: USAspending.gov was authorized by the Federal Funding Accountability and Transparency Act (2006) to provide information on the investment of U.S. tax dollars. The site provides data on federal contracts, grants, loans, and other types of spending. The website is maintained by the Office of Management and Budget.

Unit of analysis: Discrete federal grants, contracts, and loans, including sub awards

Key Data Elements: Substantial detail on each project, including dollar amount, purpose, funding agency, recipient, location of recipient, place of performance, start date, end date

Form: Database allows user-defined analysis, and visualization

Timeframe: Data are available for FY 2000 to FY 2012.

Frequency: Updates are published daily.

Access: Public; multi-screen data search, pre-formatted top picks, and FTP site flat file

Potential Uses for Economic Analysis: The federal government is a major investor in economic activity in the US. In an effort to improve government accountability, this data resource offers information on the Federal spending process. Specifically, this database provides information on government contracts, grants,

loans, direct payments, and other assistance transactions. Federal government investments account for a large portion of Identify the extent and nature of federal spending by geography over time

For Additional Information:

- Website: <http://www.USAspending.gov/>
- Email
- Phone: 800-333-4636



USPTO: U.S. Patent and Trademark Office

Category: R&D, innovation, and commercialization

Overview: The U.S. Patent and Trademark Office (USPTO) serves as the central repository for U.S. intellectual property in the form of patents and trademarks. The USPTO provides public access to a variety of statistics and datasets and is expanding its portfolio of research datasets.

Unit of Analysis: Patent and trademark filings and grants

Coverage: All U.S. patent and trademark filings and grants.

Form: Searchable and downloadable databases from USPTO and Google, and aggregate tables from USPTO

Key Data Elements:

- Patents: Patent number, application date, issue date, description, claims, inventor information (name, city and state/country), patent attorney or agent information, assignee information (name, city and state/country)
- Trademarks: Registration number, filing date, registration date, registration class, mark identification and drawing, renewal date(s) and information, registrant/assignee information

Frequency: Updated every Tuesday

Access:

- Searchable databases of published patent grants and applications and registered trademarks and applications.

- Patent and trademark public data in bulk form. Bulk data product availability is detailed here.
- Aggregate calendar year data tables of various patent statistics, including applications and grants by industry, inventor and regional areas.
- Through Google, free online access to data products that are available from USPTO on a fee basis or only on physical media. For access: bulk data.
- Google also has a special data product, Public PAIR (Patent Application Information Retrieval) data. USPTO authorizes Google to "mine" data from its Web site during hours of low usage. This arrangement serves as a bridge until such time that the USPTO is able to directly offer this data in bulk format. For access: public PAIR bulk data.

Potential Uses for Economic Analysis: Patent data offers critical information on innovative research activity. Not only does the USPTO provide data on IP, the patent record provides information on the inventor, assignee, prior art, and citations. This allows analysts not only to examine network features of the inventors but also to technological output and information flows.

For Additional Information:

- Website: USPTO
- Contact: Sandy Phetsaengam
(Duang.Phetsaengam@uspto.gov)

**ASSOCIATION OF
PUBLIC AND
LAND-GRANT
UNIVERSITIES**



University Economic Impact Metrics

Category: R&D, innovation, commercialization; business creation & development

Overview: The Association of Public and Land-grant Universities (APLU) is a research advocacy organization of public research universities, land-grant institutions, and state university systems with member campuses in 50 states and the District of Columbia. APLU's Commission on Innovation, Competitiveness and Economic Prosperity (CICEP) is developing a set of metrics that universities can use to describe their contributions to regional economies. The initiative's goal is to create a resource for universities to better measure and describe their multi-faceted contributions to innovation and economic growth.

Unit of Analysis: Universities

Coverage: Members of APLU initially. APLU also will promote the adoption of the metrics by other universities.

Size: At present, 35 academic institutions participate in the CICEP Metrics pilot. In fall 2012, APLU will encourage all 218 member institutions to adopt the metrics.

Form: To be determined

Frequency: Annual

Key Data Elements (under development):

Relationship with industry	Developing the regional and national workforce	Knowledge incubation & acceleration programs
<ul style="list-style-type: none">• Material transfer agreements• Consortia agreements• Sponsored research by industry• Clinical trials• Service to external clients	<ul style="list-style-type: none">• Student employment on funded projects• Student economic engagement• Student entrepreneurship• Alumni in the workforce	<ul style="list-style-type: none">• Incubation & acceleration program success• Relationships between client/program participants & host university• Ability to attract external investment

Data Collection Methods: Institutions primarily collect data from internal files, surveys of students and alumni, and state employment records.

Access: Through individual institutions. APLU is likely to make data available, but the means and level of aggregation has not been determined.

Potential Uses for Economic Analysis: With a focus on universities and their impact, this data source offers information for assessing individual university’s economic contributions to a region; comparing economic roles of universities in various regions; developing a typology of universities by type of contribution; and determining the implications for university policy and programs. Universities serve as a foundational base for economic activity in regions, thus it is imperative to understand how these institutions contribute to local industrial activity through education and training programs and research activity.

For Additional Information:

- Website: APLU CICEP Metrics
- Contact: Jim Woodell, Director of Innovation and Technology Policy (jwoodell@aplu.org, 202-478-6044)



Wanted Analytics

Category: Labor markets

Overview: WANTED *Analytics*TM provides real-time intelligence on labor markets. Clients use WANTED *Analytics*TM to analyze employment trends, gather competitive intelligence, forecast economic conditions, source hard-to-fill positions, and find sales leads.

Form: User-defined datasets, aggregate data tables

Unit of Analysis: Job openings

Coverage: Online job announcements

Key Data Elements: Include occupation, company, location, job competencies, educational and experience requirements

Timeframe: Collecting information since 2005

Frequency: Database is updated daily as new online job ads become available

Data Sources: Compiled from online job boards, corporate HR and government websites

Access: Subscription

Potential Uses for Economic Analysis: This source allows for a timely analysis of regional labor market characteristics and dynamics. Data on job openings and an understanding of skills that are in demand provide insights into the types of industrial activities that are expanding and currently hiring. Understanding the skills and experience requirements of firms is useful for educational planning. In addition, tracking occupational data provides an appreciation of newly emerging industrial activity as well as an understanding of how industrial structure is changing.

For Additional Information:

- Website: Wanted Analytics
- Contact: Carolyn Menz
(carolyn.menz@wantedanalytics.com)

WEB OF KNOWLEDGESM

Web of Knowledge

Categories: Networks and relationships; R&D, innovation, and commercialization

Overview: *Thomson Reuters Web of KnowledgeSM* provides access to citations for the sciences, social sciences, arts, and humanities. It is a research platform that provides access to objective content and tools to search, track, measure, and collaborate in the sciences, social sciences, arts, and humanities.

Unit of Analysis: Individual publications & academic articles

Coverage and Size: The *Web of KnowledgeSM* and *Web of ScienceSM* provide data from the following:

- 23,000 journals
- 23 million patents from 40 patent-issuing countries
- 110,000 conference proceedings
- 9,000 web sites
- 2 million chemical structures
- 87 million source items
- 700 million cited references
- 256 scientific disciplines

Form: Micro-level database; user defined table

Key Data Elements:

- Abstracts
- Citations index
- Derwent innovation index
- Science indicators

Timeframe: 100 years of back files and citation data

Frequency: Updated in real-time

Data Sources:

- Web of ScienceSM
- Chinese Science Citation Database
- Current Contents Connect
- Derwent Innovations Index
- BIOSIS Previews
- Biological Abstracts®
- CABI
- Inspec
- Medline
- Food Science and Technology Abstracts
- Zoological Record
- Journal Citation Reports®
- Essential Science Indicators

Access: Subscription for database. Free monthly online reports available on demand by occupation and metro area.

Potential Uses for Economic Analysis: Data on research-based partnerships, publications, and citations offers insight on understanding the social capital of a network, community and region. Moreover, Web of Knowledge allows analysts to identify the impacts of scientific research on the works of others with data on forward citations. Not only does this resource provide a snap shot of regional and interregional networks and relationships, this data source allows analysts to trace the trajectory of new research endeavors.

For Additional Information:

- **Website:** Web of Knowledge
- **Contacts:**
 - Matt Probus (matt.probus@thomsonreuters.com)
 - Elizabeth Deitz (elizabeth.deitz@thomsonreuters.com)



Windows Azure Marketplace—Microsoft

Category: Big data, open data platforms, and web services

Overview: The Windows Azure™ Marketplace allows users to find a wide variety of data, including demographic, environment, financial, and retail, and purchase applications to analyze those data.

Scope: The Marketplace currently provides access to 126 data sources. Many aggregate multiple data sources.

Form: Open data platform

Key Data Topics: Datasets are organized by 17 categories (e.g., business and finance) and by publisher.

Access: Free, free trial, or subscription, depends on dataset

Potential Uses for Economic Analysis: Windows Azure offers access to a large number of datasets relevant to regional analysis. This application platform provides a cyber-infrastructure that allows researchers to integrate their public cloud applications in their existing IT environment.

For Additional Information:

- Website: Windows Azure Marketplace-Microsoft
- Contact: Avi Kovarsky (avi.kovarsky@microsoft.com)



YourEconomy.org

Categories: Business creation and development; labor markets; longitudinal databases; regional industries and economies

Overview: YourEconomy.org (YE) is a user-friendly site that draws on the annual National Establishment Time Series (NETS) to provide business census statistics. YE captures regional growth, at state, county, and MSA-levels, in terms of employment, sales, and establishment dynamics. Results are filterable by 3-digit NAICS codes. In YE calculations, local businesses, local businesses with non-local headquarters, and non-business employers are included. The tool is freely available, with more detailed data available as part of a premium subscription.

Units of Analysis: Establishments, jobs, and sales

Coverage: Firms in the Dun & Bradstreet database

Level of Detail: 3-digit NAICS code, MSA and county level

Form: Interactive inquiry

Key Data Elements:

- Growth factors, resulting in establishment & employment gains and losses
 - Births/deaths
 - Expansions
 - Relocations
- Establishment size
- Number of establishments within a defined geography
- Metrics are divided by sector type
 - Noncommercial (educational institutions, government, nonprofits, etc.)
 - Nonresident (businesses with non-local headquarters)
 - Resident (local businesses)

Size: 25 million active business establishments; 44 million total establishments over time

Timeframe: 1990 to present (with 12 to 18-month lag).

Frequency: Annual

Data Sources: From the National Establishment Time Series (NETS); tracks firms using their assigned Dun and Bradstreet Number (DUNS).

Access: General access free; Premium service provides detailed job and establishment data, including openings, closings, expansions, contractions, and relocations

Potential Uses for Economic Analysis: YE provides data to identify the dynamics of regional economic change with a focus on real-time establishment and job activity. Specifically, this source offers data on the expansions of existing establishments, impacts of relocating firms, and job contributions from resident and non-resident establishments.

For Additional Information:

- Website: YourEconomy.org
- Contact: T.J. Becker, Marketing and Media
 - Phone: (269) 445-4294
 - Email: tjbecker@lowe.org

Conference Attendees

AEM Corporation

Baron Rodriguez

Director of Strategic Education Initiatives

baron.rodriguez@aemcorp.com

Akundi Inc

Krishna Akundi

Analyst

ag.krishnamohan@gmail.com

American Association of Community Colleges

Rahel Tekle

Research Associate

rtekle@aacc.nche.edu

American Economic Association

Daniel Newlon

Director, Government Relations

dan.newlon@aeapubs.org

American Express

Ravi Varma

Vice President- Risk & Information Management

ravi.v.varma@aexp.com

American Statistical Association

Steve Pierson

Director of Science Policy

pierson@amstat.org

Artemel & Associates Inc.

Agnes Artemel

President

apa@artemel.com

Association for Career and Technical Education

Catherine Imperatore

Research Manager

cimperatore@acteonline.org

Association of American Medical Colleges

Lutheria Peters

Research Analyst

lpeters@aamc.org

Association of American Universities

Josh Trapani

Senior Policy Analyst

josh_trapani@aau.edu

Association of Public and Land-Grant Universities

Jim Woodell

Director of Innovation and Technology Policy

jwoodell@aplu.org

AVIAN Engineering

Beth Masimore

Technical Analyst

beth.a.masimore@ugov.gov

Baltimore City Mayor's Office of Employment Development

Milena Kornyl

Senior Research and Evaluation Specialist

mkornyl@oedworks.com

BForlaw, LLC

Blair Forlaw

Economic Development Consultant

blair@bforlaw.com

Board of Governors of the Federal Reserve System

Barbara Robles

Senior Research Liaison

barbara.j.robles@frb.gov

Brookings Institution

Alec Friedhoff

Data Manager & Research Analyst, Metropolitan Policy Program

afriedhoff@brookings.edu

Martha Ross

Fellow, Metropolitan Policy Program

mross@brookings.edu

Jonathan Rothwell

Associate Fellow

jrothwell@brookings.edu

Bureau of Economic Analysis, U.S. Department of Commerce

Zoe Ambargis

Regional Economist

zoe.ambargis@bea.gov

Terence Fallon

Economist

terence.fallon@bea.gov

Christopher Lucas

Economist

Christopher.Lucas@bea.gov

Brian Maisano

Economist

Brian.Maisano@bea.gov

Nathaniel Milhous

Economist

Nathaniel.Milhous@bea.gov

Sharon Panek
Section Chief
sharon.panek@bea.gov

Carol Robbins
Chief, Analysis and Special Studies Branch, Region
carol.robbsins@bea.gov

Matthew von Kerczek
Economist
Matthew.vonkerczek@bea.gov

Bureau of Labor Statistics, U.S. Department of Labor
Akbar Sadeghi
Economist
sadeghi.akbar@bls.gov

Patrick Carey
Branch Chief, Mass Layoff Statistics
carey_p@bls.gov

Michael LoBue
Economist
lobue.michael@bls.gov

Sandra Mason
Division Chief, Local Area Unemployment Statistics
mason.sandi@bls.gov

Walter Sylva
Economist
sylva.walter@bls.gov

Robert Viégas
Economist
viegas.robert@bls.gov

Capital Area Council of Governments

John Rees

Director of Economic Development

jrees@capcog.org

Census Bureau, U.S. Department of Commerce

Shirin Ahmed

Assistant Associate Director of Economic Programs

shirin.a.ahmed@census.gov

Dierdre Bevington-Attardi

Geographer

dierdre.bevington.attardi@census.gov

Lisa Blumerman

Chief, Governments Division

Lisa.m.blumerman@census.gov

William Bostic

Associate Director for Economic Programs

william.g.bostic.jr@census.gov

Cavan Capps

Senior Technology Advisor, Research & Methodology

cavan.paul.capps@census.gov

Rick Denby

Assistant Division Chief

richard.a.denby@census.gov

Carma Hogue

Assistant Division Chief, Governments Division

carma.ray.hogue@census.gov

Richard Hough

Branch Chief

richard.s.hough@census.gov

Ron Jarmin
Assistant Director for Research and Methodology
ron.s.jarmin@census.gov

Bob Marske
Chief, Product Development Staff
robert.a.marske@census.gov

Nancy Potok
Associate Director for Demographic Programs
nancy.a.potok@census.gov

Mark Wallace
Division Chief
mark.e.wallace@census.gov

Center for American Progress
Sean Pool
Managing Editor, Science Progress
spool@americanprogress.org

Center for Economic Development, Cleveland State University
Iryna Lendel
Assistant Director
i.lendel@csuohio.edu

Center for International Strategy, Technology, and Policy, Georgia
Institute of Technology
Dae Yeon Cho
Visiting Scholar
daeyeoncho08@gmail.com

Center for Real Estate & Economic Development, Old Dominion
University
John Lombard
Director
jlombard@odu.edu

Center for Regional Analysis, George Mason University

Ryan Price

Research Associate

ryantprice@gmail.com

Center for Regional Economic Competitiveness

Lauren Gilchrist

Research Analyst

lgilchrist@crec.net

Sarah Gutschow

Research Assistant

sgutschow@crec.net

Ronald Kelly

Director, Workforce Research and Training

rkelly@crec.net

Center for Science, Technology & Economic Development, SRI
International

Jeffrey Alexander

Senior Policy Analyst

jeffrey.alexander@sri.com

Christina Freyman

Policy Analyst

christina.freyman@sri.com

Center on Budget and Policy Priorities

Frank Mamo

Research Intern

06frmamo@gmail.com

Vincent Palacios

Research Associate

palacios@cbpp.org

Hannah Shaw
Research Associate
shaw@cbpp.org

Center on Wisconsin Strategy (COWS)

Michele Mackey
Senior Associate
mmackey@cows.org

City University of New York

Punit Arora
Assistant Professor
punitsarora@gmail.com

Cleveland State University

Candi Clouse
Research Associate
c.clouse@csuohio.edu

Coalition of Urban Serving Universities/Association of Public and
Land Grant Universities

Shari Garmise
Vice President
sgarmise@aplu.org

Columbia University

Philipp Brandt
Researcher
psb2120@columbia.edu

Complete College America

Mike Baumgartner
Vice President, Finance and Special Projects
mbaumgartner@completecollege.org

Congressional Research Service

Laura Hanson

Information Research Specialist

lhanson@crs.loc.gov

CONNECT

Timothy Tardibono

VP of Public Policy

timothy@connect.org

Consultant

Sharon Brown

Consultant

sharonpbrown@verizon.net

Corporation for Enterprise Development (CFED)

Anita Drever

Director of Applied Research

adrever@cfed.org

Michelle Nguyen

Senior Research Associate

mnguyen@cfed.org

Council of Economic Advisers

Adrienne Pilot

Director of Statistical Office

apilot@cea.eop.gov

Council of Professional Associations on Federal Statistics

Edward Spar

Executive Director

copafs@aol.com

DC Department of Employment Services

Sania Khan

Labor Economist

sania22@gmail.com

Miranda Logan
Labor Market Information Specialist
miranda.logan@dc.gov

Rebecca Royster
Labor Market Specialist
rebecca.royster@dc.gov

DC Department of Employment Services, Office of Policy,
Performance & Economics

James Moore, Jr
Deputy Director
james.moore4@dc.gov

DC Department of Employment Statistics

John Kangethe
Labor Economist, Labor Market Information
john.kangethe@dc.gov

Dell SecureWorks

Raj Bandyopadhyay
Software Engineer
rbandyopadhyay@secureworks.com

Economic Development Administration, U.S. Department of
Commerce

Bryan Borlik
Director, Performance and National Programs Division
Bryan.borlik@eda.gov

Thomas Guevara
Deputy Assistant Secretary
thomas.christopher.guevara@eda.gov

Casey Muhm
Program Analyst

casey.j.muhm@eda.gov

Hillary Sherman
Program Analyst - Team Lead
hillary.m.sherman@eda.gov

Economic Modeling Specialists Inc
Andrew Crapuchettes
CEO
prench@economicmodeling.com

Economic Research Service, U.S. Department of Agriculture
Sarah Low
Economist
slow@ers.usda.gov

Andrew Toole
Research Economist
atoole@ers.usda.gov

Economics and Statistics Administration, U.S. Department of Commerce
Brittany Bond
Economist
bbond@doc.gov

Sandra Cooke-Hull
Economist
scookehull@doc.gov

Cassandra Ingram
Economist
cingram@doc.gov

Electric Fork Media
Tracy Bartley
Visual Information Specialist
Tracyebartley@gmail.com

Elsevier

Cynthia Cleto

Consultant

c.cleto@elsevier.com

Steve Lagerveld

Account Manager

s.lagerveld@elsevier.com

Eric Livingston

SciVal Solutions Manager

E.Livingston@Elsevier.com

Josine Stallinga

Market Development Manager

j.stallinga@elsevier.com

Employment and Training Administration, U.S. Department of
Labor

Dais Anthony

Workforce Information Program Manager

dais.anthony@dol.gov

Kate Louton

Analyst

louton.kathryn@dol.gov

Engineering Systems Division, Massachusetts Institute of
Technology

Liliana Rivera

Student ESD-MIT

mlrivera@mit.edu

EUROREG, University of Warsaw

Karol Olejniczak

Assistant Professor

k.olejniczak@uw.edu.pl

Federal Reserve Bank of Boston

Robert Clifford

Policy Analyst

robert.clifford@bos.frb.org

Federal Reserve Bank of Richmond

Andy Bauer

Regional Economist

andy.bauer@rich.frb.org

Federal Reserve Board

David Byrne

Economist

david.m.byrne@frb.gov

Anna Mkhitarian

Community Development Intern

anna.mkhitarian@frb.gov

Federal Reserve Board of Governors

Clifford Follweiler

Senior Analyst

cliffordross@hotmail.com

Federal Reserve System

Mario Arthur-Bentil

Research Assistant

m.arthurbentil@gmail.com

Gates Foundation

Nick Lee

Senior Program Officer

Nick.lee@gatesfoundation.org

George Mason University

Darrene Hackler

Associate Professor

dhackler@gmu.edu

George Washington Institute of Public Policy

Patricia Atkins

Associate Research Professor

patkins@gwu.edu

Kari Nelson

Research Assistant

knelson9@gwu.edu

Eric Stokan

Research Assistant

ericjstokan@gmail.com

George Washington University

Stephanie Cellini

Assistant Professor

scellini@gwu.edu

Sarah Ficenec

PhD Student

sficenec@gmail.com

James Foster

Professor of International Affairs and Economics

fosterje@gwu.edu

Tom Harrod

Librarian

tph@gwu.edu

Georgetown University Center on Education and the Workforce

Andrew Hanson

Research Analyst

andrew.hanson@me.com

Jeff Strohl

Director of Research

js787@georgetown.edu

Georgia Institute of Technology

Usha Nair-Reichert

Interim Chair

usha.nair@gatech.edu

Government Accountability Office

Ron Fecso

Chief Statistician

fecsor@gao.gov

IDA Science and Technology Policy Institute

Daniel Basco

Fellow

dbasco@ida.org

Stephanie Shipp

Research Staff Member

sship@ida.org

IHS Global Insight

Tasmina Ahmed

Analyst

Tasmina.Ahmed@ihs.com

Tom Jackson

Senior Economist

tom.jackson@ihs.com

Michael Lynch
Senior Economist
michael.lynch@ihs.com

David Payne
Senior Economist
dpayne@doc.gov

Illinois Science & Technology Coalition
Alya Adamany
Director, Innovation and Special Initiatives
aadamany@istcoalition.org

ILR School, Cornell University
Diane Burton
Associate Professor
burton@cornell.edu

Indiana Business Research Center, Indiana University
Jerry Conover
Director
conover@indiana.edu

Timothy Slaper
Research Director
tslaper@indiana.edu

Information Technology & Innovation Foundation (ITIF)
Luke Stewart
Economic Analyst
lstewart@itif.org

Innovation Ecologies Inc
Robin Gaster
President
rgaster5@gmail.com

Jacob France Institute, University of Baltimore

Seema Iyer

Associate Director

siyer@ubalt.edu

Jing Li

Research Analyst

jli@ubalt.edu

Ting Zhang

Assistant Professor

tzhang@ubalt.edu

Jobs for the Future

Robert Holm

Program Director

rholm@jff.org

Myriam Milfort

Senior Project Manager

mmilfort@jff.org

Todd Weissman

Senior Project Manager

tweissman@jff.org

John Glenn School of Public Affairs

David Landsbergen

Associate Professor

landsbergen.1@osu.edu

Johns Hopkins University

Cathy Lebo

Assistant Provost

lebo@jhu.edu

Susan Morris
Fellow
s.morris@jhu.edu

LEEDS School of Business, University of Colorado
Siddharth Vedula
PhD Student
siddharth.vedula@colorado.edu

Library of Congress
Julie Jennings
Information Research Specialist
jjennings@crs.loc.gov

Loudoun County Department of Economic Development
Olga Deyneko
Industry Research Analyst
olga.deyneko@loudoun.gov

Donna Hurwitt
Business Research Analyst
donna.hurwitt@loudoun.gov

Market Street
Ellen Cutter
Director of Research
ecutter@marketstreetservices.com

Maryland Department of Business and Economic Development
Nancy McCrea
Research Director
nmccrea@choosemaryland.org

James Palma
Senior Manager, Research and Information
jpalma@choosemaryland.org

Michigan State University

Mark Largent

Associate Professor

largent@msu.edu

Mission de Reflexion sur la Consolidation de la Democratie au
Mali (MRDCM)

Moussa Konte

Assistant Juridique

soupleya@yahoo.fr

N/A

Ed Gilliland

ed.gilliland@verizon.net

National Academy of Sciences

Connie Citro

Director

ccitro@nas.edu

Daniel Cork

Senior Program Officer

dcork@nas.edu

Ana Ferreras

Senior Program Officer

aferreras@nas.edu

Hermann Habermann

Senior Program Officer

hhabermann@nas.edu

Chris Mackie

Senior Program Officer

cmackie@nas.edu

National Association of Development Organizations

Robert Black
Program Manager
mblack@nado.org

Carrie Kissel
Associate Director
ckissel@nado.org

National Association of State Directors of Career Technical
Education Consortium

Kara Herbertson
Education Policy Analyst
kherbertson@careertech.org

National Center for Science and Engineering Statistics, National
Science Foundation

John Jankowski
Program Director
jjankows@nsf.gov

Nirmala Kannankutty
Senior Advisor
nkannank@nsf.gov

Christopher Pece
Senior Analyst
cpece@nsf.gov

National Credit Union Administration

Benjamin Lidofsky
Economist
benjamin.lidofsky@uvm.edu

National Education Association

Richard Sims
Chief Economist
rsims@nea.org

National Endowment for the Arts

Bonnie Nichols
Research Analyst
nicholsb@arts.gov

National Institute of Food and Agriculture

Charlotte Baer
National Program Leader for Animal Science
cbaer@nifa.usda.gov

Robbin Shoemaker
National Program Leader for Economics
rshoemaker@nifa.usda.gov

National Institute of Mental Health, National Institutes of Health

Liza Bundesen
Health Scientist Administrator
lbundese@mail.nih.gov

National League of Cities

Katie McConnell
Senior Associate, Economic Development
mcconnell@nlc.org

Christy McFarland
Program Director, Finance and Economic Development, Center for
Research and Innovation
Mcfarland@nlc.org

National Research Council

Kaye Husbands Fealing
Professor/Study Director
kfealing@nas.edu

National Science Board

John Veysey
Science Policy Analyst

jveysey@nsf.gov

National Science Foundation

David Croson

Program Director

dcroson@nsf.gov

Lee Herring

Speech Writer

lherring@nsf.gov

Rebecca Rosen

AAAS Policy Fellow

rrosen@nsf.gov

C Renee Wilkerson

Program Analyst

cwilkers@nsf.gov

Raymond Wolfe

Economist

rwolfe@nsf.gov

National Skills Coalition

Andy Van Kleunen

Executive Director

andyvk@nationalskillscoalition.org

North Carolina Rural Economic Development Center

John Killeen

Senior Associate for Statistical Analysis

jkilleen@ncruralcenter.org

Odum School of Ecology, University of Georgia

Rebeca De Jesus-Crespo

PhD Research Assistant

rdejesus@uga.edu

Office of Economic Adjustment, U.S. Department of Defense

Michael Gilroy
Project Manager
michael.gilroy@wso.whs.mil

Margit Myers
Project Manager
margit.myers@wso.whs.mil

Liz Chimienti
Presidential Management Fellow
elizabeth.chimienti@wso.whs.mil

Office of Innovation and Entrepreneurship, Economic
Development Administration, U.S. Department of Commerce
Paul Corson
Deputy Director
pcorson@doc.gov

Office of Management and Budget
Aaron Ampaw
Program Examiner
aampaw@omb.eop.gov

Michael Branson
Program Examiner
mbranson@omb.eop.gov

Paul Bugg
Economist
pbugg@omb.eop.gov

Rachel Zinn
Program Examiner
rzinn@omb.eop.gov

Office of Research Information Systems, National Institutes of
Health

Israel Lederhendler
Senior Science Advisor
lederhei@od.nih.gov

Office of Science and Technology Policy, Executive Office of the President

Kei Koizumi
Assistant Director Federal Research & Development
Kei_Koizumi@ostp.eop.gov

Office of the Chief Economist, U.S. Department of Commerce

David Beede
Economist
dbeede@doc.gov

Ohio State University

Jenny Brodie
PhD Student/Policy Analyst
brodie.7@osu.edu

Ontario Ministry of Finance

Tatiana Muntean
Senior Economist, Office of Economic Policy
tatiana.muntean@ontario.ca

Optimal Solutions Group, LLC

Berna Demiralp
Senior Research Associate
bdemiralp@optimalsolutionsgroup.com

Mark Turner
President and Chief of Optimal Solutions Group LLC
mturner@optimalsolutionsgroup.com

Pew Center on the States, Pew Charitable Trusts

Brandon Brockmyer
Senior Research Analyst, Research & Information
bbrockmyer@pewtrusts.org

Jeff Chapman
Manager
jchapman@pewtrusts.org

Ike Emejuru
Associate
Emejuruik@yahoo.com

Ricky Harrison
Senior Research Associate
rharrison@gmail.com

Julie Holzhuetter
Associate
jholzhuetter@pewtrusts.org

Sergio Ritacco
Leadership Year Fellow, Research & Information
sritacco@pewtrusts.org

Progressive Policy Institute
Michael Mandel
Chief Economic Strategist
mmandel@ppionline.org

Public Health Law Research
Heidi Grunwald
Deputy Director
grunwald@temple.edu

Regional Research Institute, West Virginia University
Gianfranco Piras
Research Assistant Professor
Gianfranco.Piras@mail.wvu.edu

Regional Technology Strategies Inc.
Chris Beacham

Principal
cbeacham@rtsinc.org

Ross School of Business, University of Michigan

Bo Zhao
Doctoral Candidate
bozhao@umich.edu

RTI International

Sara Lawrence
Senior Economic Development Analyst
slawrence@rti.org

Rutgers Business School, Rutgers University

Brett Anitra Gilbert
Assistant Professor
bagilbert@business.rutgers.edu

Rutgers University

Quintus Jett
Assistant Professor
qjett@andromeda.rutgers.edu

RW Ventures

Sophie Cohen
Associate
sophie@rw-ventures.com

School of Business, George Washington University

Paul D Reynolds
Research Professor of Management
pauldavidsonreynolds@gmail.com

School of City and Regional Planning, Georgia Institute of Technology

Nathanael Hoelzel
PhD student
natehoelzel@gatech.edu

Nancey Green Leigh
Professor
ngleigh@gatech.edu

School of Public Policy, George Mason University
Nobuhiko Daito
PhD Student / Graduate Research Assistant
ndaito@gmu.edu

Yu Jin Jung
PhD Candidate
yjung4@gmu.edu

School of Public Policy, Georgia Institute of Technology
Kirsten Bandyopadhyay
Doctoral Student
kirstenb@gatech.edu

Jennifer Clark
Associate Professor
jennifer.clark@gatech.edu

School of Regional Planning, Georgia Institute of Technology
Thomas Douthat
PhD Research Assistant
tdouthat@gatech.edu

Science and Technology Policy Institute
Elaine Sedenberg
Baccalaureate Fellow
esedenbe@ida.org

Elmer Yglesias
Research Associate
eyglesia@ida.org

Sole Proprietor

Alicia Giudice
Land Use Planner
aligiudice@yahoo.com

Southern California Association of Governments

Cheol-Ho Lee
Senior Regional Planner
leec@scag.ca.gov

Southern Rural Development Center

Roberto Gallardo
Assistant Extension Professor
robertog@srdc.msstate.edu

SRI International

John Chase
Senior Analyst
john.chase@sri.com

Startup America Partnership

Kathleen Warner
Chief Operating Office
kathleen@s.co

SYNEVA Economics LLC

Tom Tveidt
Research Economist
tom@syneva.com

Synthosys

Ronald Lai
Business Consultant
laironald@gmail.com

Peter Rosenthal
Software Engineer
peter.rosenthal@synthosys.com

The Evidence Network

Brian Barge

President and CEO

barge@theevidencenetwork.com

Margaret Dalziel

VP Research

dalziel@theevidencenetwork.com

The Global Impact Investing Network

Kelly McCarthy

Manager, IRIS Adoption

kmccarthy@thegiin.org

The MITRE Corporation

Stephen Stratman

Artificial Intelligence Engineer

sstratman@mitre.org

TischlerBise

Julie Herlands

Principal

julie@tischlerbise.com

Total-Impact

Jason Priem

Doctoral Student

jp@jasonpriem.org

Trachtenberg School of Public Policy and Public Administration,

George Washington University

Joseph Cordes

Professor of Economics, Public Policy and Public Admin

cordes@gwu.edu

Twin Ravens Consulting

Brandon Shackelford

Owner

brandon@twinravensconsulting.com

U.S. Department of Commerce

Christian Awuku-Budu

Economist

christian.awuku-budu@bea.gov

Derek Ramirez

Economist

dramirez@nccommerce.com

Duke Tran

Regional Economist

duke.tran@bea.gov

U.S. Department of Education

Anthony Kelly

Education Research

anthony.kelly@ed.gov

U.S. Department of Housing and Urban Development

Josh Geyer

Analyst

Joshua.M.Geyer@hud.gov

Sarah Schaefer

Social Science Analyst

sarah.k.schaefer@hud.gov

U.S. Government Accountability Office

William Shear

Director, Financial Markets and Community Investment

shearw@gao.gov

U.S. Green Building Council

Andre Poremski

Creative Director

aporemski@usgbc.org

U.S. International Trade Commission

Samira Salem

Program Manager for Labor

samira.salem@usitc.gov

U.S. Securities & Exchange Commission

Jonathan Karp

Division Information Officer

karpj@sec.gov

University at Albany

Laura Schultz

Assistant Professor

lschultz@albany.edu

University of California- San Diego

Josh Shapiro

Director of Research and Evaluation

jshapiro@ucsd.edu

University of Hamburg/UNC

Max-Peter Menzel

Professor

menzel@geowiss.uni-hamburg.de

University of Illinois at Urbana-Champaign

Janet Bercovitz

Associate Professor

jbercov@illinois.edu

University of Maryland

Scott Dempwolf

PhD Candidate

dempy@umd.edu

Kenneth Durril

Academic Advisor

kdurril@umd.edu

Brent Goldfarb
Associate Professor of Entrepreneurship
brent.goldfarb@gmail.com

Miguel Sarzosa
PhD Student
miguel_sarzosa@hotmail.com

University of Miami
Richard Bookman
Senior Advisor to the Dean
rbookman@miami.edu

University of Michigan
Jason Owen-Smith
Associate Professor, Sociology & Organizational St
jdos@umich.edu

University of North Carolina, Chapel Hill
Nichola Lowe
Associate Professor
nlowe@unc.edu

University of Washington
Shawn Medero
Platform Program Lead, Office of Research Information Services
smedero@uw.com

Urban Institute
Chris Narducci
Research Associate
cnarducci@urban.org

Virginia Economic Development Partnership
Charles Kennington
Economist
ckennington@yesvirginia.org

Workforce Enterprise Services, Inc.

Timothy Harmon

Consultant

tharmon@goworkforce.com

World Business Chicago

Liz Jellema

Director of Research

ljellema@worldbusinesschicago.com

Symposium: Use of Innovative Data Academic Conference

Use of Innovative Data Sets for Regional Economic Research

May 9, 2012

9:00AM – 5:00PM

Marvin Center, Room 309

800 21st Street, NW Washington, DC 20052

Agenda

9:00 – 9:05am: Welcome and Introduction

- Hal Wolman, GWIPP Director

9:05 – 9:45am: Labor Force, Moderator: Hal Wolman;

Discussant: Paul Reynolds

- Erika McEntarfer, Center for Economic Studies, U.S. Census Bureau, “Job-to-Job Flows and the Business Cycle”

9:45 – 10:30am: University R&D, Moderator: Hal Wolman;

Discussant: Paul Reynolds

- Jason Owen-Smith, University of Michigan, “Doing Policy Relevant Research Using StarMetrics”
- Shanu Sushmita, UCLA, “Moving Discoveries From Science to Commerce”

10:30 – 10:45am: Break

- Coffee, pastries

10:45am – 12:15pm: Companies, Moderator: Andrew Reamer;

Discussant: Paul Reynolds

- Martin Kenney, University of California – Davis, “Start-Ups, Employment Growth, and Geography: Results from Emerging Growth Company IPOs, 1996-2010.”

- Don Walls, Walls & Associates, “Which Metropolitan Markets Are Best at Fostering New Firms that Survive?”
- Maryann Feldman and Nichola Lowe, University of North Carolina at Chapel Hill, “Circling the Triangle: Using a Spatial Longitudinal Company Data Base to Research the Research Triangle”
- Diane Burton, Cornell University, “Entrepreneurial Firm Development in Silicon Valley: The Use of Career History Data”

12:15 – 1:15pm: Lunch

1:15 –2:30pm: Patents, Moderator: Andrew Reamer; Discussant: Edward Feser

- David Wolfe, University of Toronto, “The Spatial Scale of Innovation”
- Dieter Kogler, University College Dublin, “Inter-Organizational Knowledge Spillovers in the Evolution of Biotechnology Invention, 1981-2010”
- Vette Torvik, University of Illinois at Urbana-Champaign, “A Dataset of Biomedical Author-Inventors: Probabilistic Disambiguation and Linking Names Across PubMed and USPTO”

2:30 – 3:15pm: Regional Industries, Moderator: Maryann Feldman; Discussant: Edward Feser

- Mercedes Delgado, Temple University, and Christian Ketels, Harvard Business School, “Clusters and Economic Performance: Leveraging the Cluster Mapping Database”
- Sharmistha Bagchi-Sen, University of Buffalo, “Researching Firms Using Primary and Secondary Data: A Case of Regional Bioenergy in the United States”

3:15 – 3:30pm: Break

- Coffee, cookies

3:30 – 4:15pm: Regional Economies, Moderator: Maryann Feldman; Discussant: Edward Feser

- David Rigby, UCLA, “Disentangling the Local and Regional Impacts of Globalization and Trade Using Census Bureau Micro-Data”
- Deborah Strumsky, University of North Carolina at Charlotte, “You Can’t Get There From Here: Movement in Metropolitan Inventive Spaces”

4:15 – 5:00pm: Conclusion

Discussants

- Paul Reynolds, The George Washington University, morning panels
- Edward Feser, University of Manchester, afternoon panels

Common Abbreviations

ACS: American Community Survey

APLU: Association of Public and Land-Grant Universities

AUTM: Association of University Technology Managers

AWS: Amazon Web Services

BDS: Business Dynamics Statistics

BEA: Bureau of Economic Analysis

BED: Business Employment Dynamics

BLS: Bureau of Labor Statistics

BRDIS: Business R&D and Innovation Survey

CES: Center for Economic Studies (U.S. Census)

CICEP: Commission on Innovation, Competitiveness and
Economic Prosperity

D&B: Dun and Bradstreet

DUNS: Dun & Bradstreet Number

EDA: Economic Development Agency, Department of Commerce

EDGAR: Electronic Data, Gathering and Retrieval

EMSI: Economic Modeling Specialists Incorporated

FTP: File transfer protocol

GDP: Gross Domestic Product

GGs: Green Goods Survey

IDEA: Indicator Database for Economics Analysis

IP: Intellectual Property

IPO: Initial Public Offering

ITWG: Interagency Technical Working Group

LBD: Longitudinal Business Database

LED: Local Employment Dynamics

LEHD: Longitudinal Employer-Household Dynamics

M&A: Merger and acquisition

MSA: Metropolitan Statistical Area

NAICS: North American Industry Classification System

NCS: National Student Clearinghouse

NETS: National Establishment Time-Series

O*NET: Occupational Information Network

OMB: Office of Management and Budget

PROGRIS: Program on Globalization & Regional Innovation
Systems

PUMS: Public Use Micro data Sample

QCEW: Quarterly Census of Employment and Wages

QWI: Quarterly Workforce Indicators

R&D: Research and Development

RDC: Research Data Centers (U.S. Census)

RLI: Real-time Labor Intelligence

RPP: Regional Price Parities

RTP: Research Triangle Park (NC)

S&P: Standard and Poor's

SaaS: Software as a Service

Sci²: Science of Science

SEC: Securities and Exchange Commission

SOC: Standard Occupation Classification

STAR METRICS: Science and Technology for America's

Reinvestment: Measuring the Effects of Research on
Innovation

STAT: Statistics Access for Tech Transfer

TEN: The Evidence Network

USITC: U.S. International Trade Commission

USPTO: U.S. Patent and Trademark Office

Common Definitions of Terms

Agglomeration: A geographic concentration of people and/or activities

API (application programming interface): A language and message format used by an application program to communicate with the operating system or some other control program

Backward linkages: Linkages to suppliers of inputs (as different from forward linkages to customers of outputs) part of economic interdependence system; useful concept to differentiate direction of flows in complex economies.

Cloud Computing: using the Internet and remote servers to maintain information and applications.

Cluster: geographic concentration of related firms in an industry

CMSA: Consolidated Metropolitan Statistical Area; unit of geographic data for the Census Bureau to describe a city and its surrounding area.

Company Capacity: Captures metrics the gauge the quality and health of the general population

Cross-sectional data: a data set of a sample of the population at one point in time

Economic development: Advancement in a region's economic health and quality of life

Economic growth: Increase in the capacity of an economy to produce goods and services

Entrepreneurial firm: New business venture

Establishment: A single physical location where business is conducted or where services or industrial operations are performed.

Firm Capacity: Extends beyond the initial formation of a firm and encompasses characteristics of the firm relevant to its own growth and survival

Founder: An entrepreneur, or person who starts a company

Intellectual property: property right to intangible assets, which include patents, trade secrets, and trademarks

Knowledge spillover: benefit from an idea or information that does not occur through a market transaction.

Liquidity events: Initial Public Offerings (IPOs), Mergers and Acquisitions (M&A); Bankruptcy

Longitudinal data: A dataset that tracks the same information on the same sample at multiple points in time.

Micro-level data: Data set of individual or household-level data gathered from surveys

MSA: Metropolitan Statistical Area; unit of geographic data for the Census Bureau

Open source application: Applications that allow public users to use and change its source code.

Place: A particular geography to which a group of people has become attached, endowing it with meaning and significance. Often associated with notions of family, home and community

R&D: Research and Development; the creative process that generates new knowledge and innovation

Regional Innovative Capacity: Considers the greater context of the innovation system; how firms relate to the larger innovative infrastructure

Technology transfer: The process of transferring skills, knowledge, and technologies among institutions such as governments and universities.

Text-mining: Extracting information from text based on pre-defined word patterns

Time Series Data: Data on an event at regular time intervals, analogous to longitudinal or panel data

Tool (data tool): Computational interface that allows users to manage and administer the content of a data source

User-defined analysis: Data sources with a greater degree of flexibility to allow researchers to custom tailor the contents and geographic dimensions of the dataset

Web archive: Saving the pages from Web sites as they change over time for historical purposes.