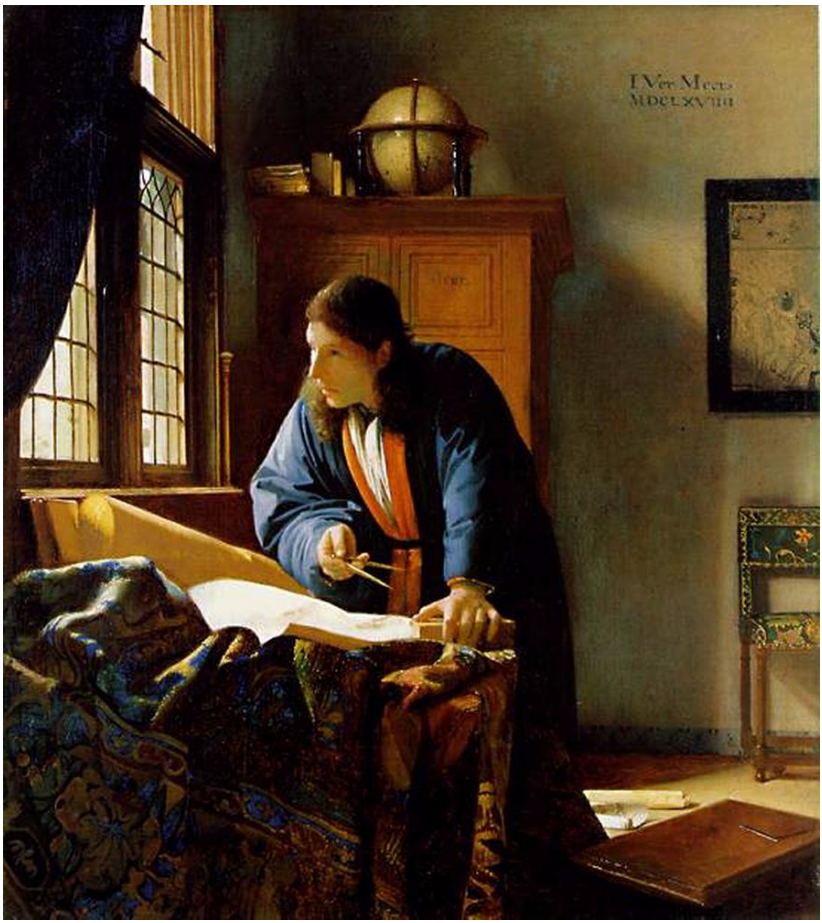
INNOVATIVE DATA SOURCES FOR REGIONAL ECONOMIC ANALYSIS



Maryann Feldman, Alex Graddy Reed, Lauren Lanahan,

Glen McLaurin, Kari Nelson, and Andrew Reamer

INNOVATIVE DATA SOURCES FOR REGIONAL ECONOMIC ANALYSIS

Conference Guide

George Washington University

Washington, DC

May 7–8, 2012

Maryann Feldman

Alex Graddy-Reed

Lauren Lanahan

Glenn McLaurin

Kari Nelson

Andrew Reamer

Acknowledgements: Funding for this project from 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 REGIONAL 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

Data Sources by Category

Big data, open data platforms, web services

Amazon Web Services (C4)

DATA.GOV (C2)

Factual (C5)

LED & LEHD: Local Employment Dynamics (C3)

PriceStats (B9)

Windows Azure Marketplace—Microsoft (C1)

Business creation and development:

University Economic Impact Metrics (C8)

BDS: Business Dynamics Statistics (A12)

BED: Business Employment Dynamics (B17)

Circling the Research Triangle (C16)

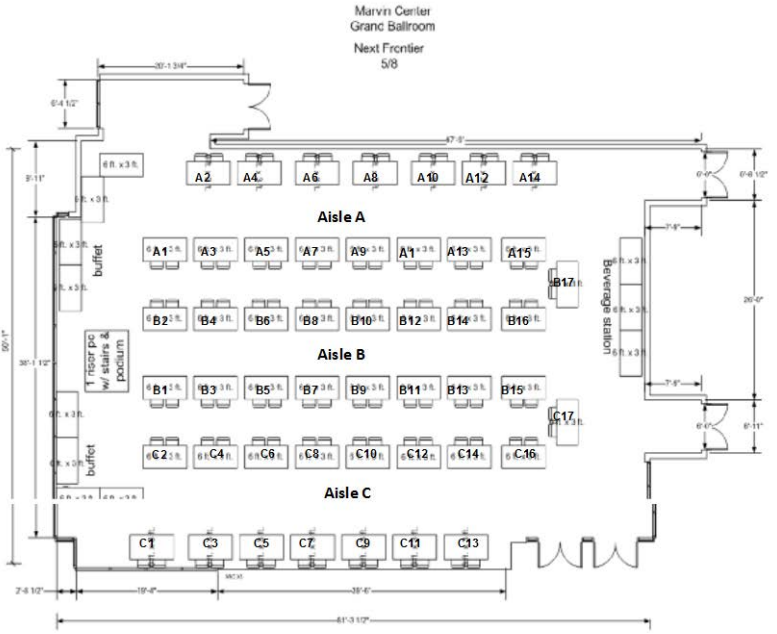
Kenny-Patton IPO Database (C5)

RDC: Research Data Centers (A11)

S&P Capital IQ (B14)

STAR METRICS (C13)

YourEconomy.org (A7)


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
B2
B4
B6
B8
B10
B12
B14
B16
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
C2
C4
C6
C8
C10
C12
C14
C16
Aisle C
C1
C3
C5
C7
C9
C11
C13
buffet
buffet
T riser pc
w/ stairs &
podium
beverage station

Data intermediaries and integrators:

DataWeb for the DataFerrett (A8)

Moody's Analytics **(**C12)

STATS America (A9)

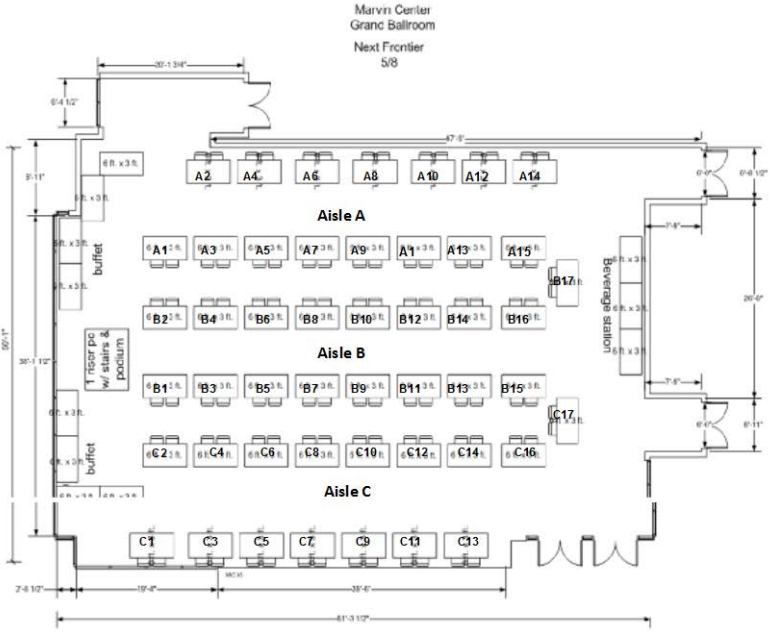
Data Analysis & Visualization Tools:

Altmetrics: Total-Impact (C6)

EMSI(A2)

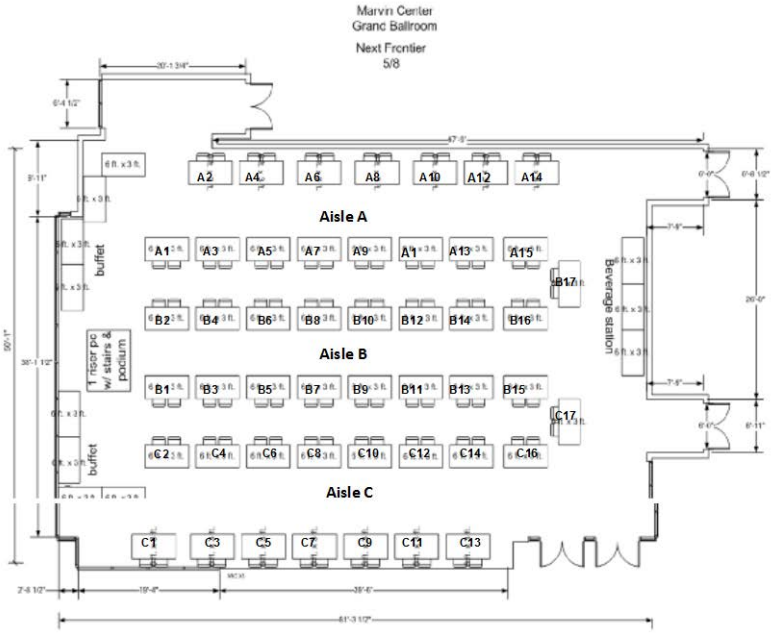
GeoIQ (A4)

Sci² (A6)


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
buffet
Beverage station
B2
B4
B6
B8
B10
B12
B14
B16
1 riser pc w/ stairs & podium
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
C2
C4
C6
C8
C10
C12
C14
C16
buffet
Aisle C
C1
C3
C5
C7
C9
C11
C13

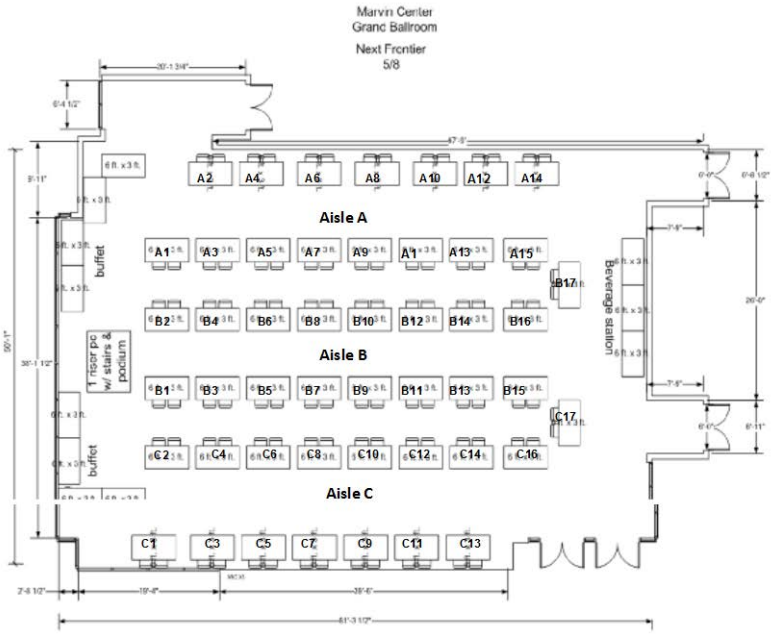
Jobs, workforce, education & labor markets:

Longitudinal Databases:

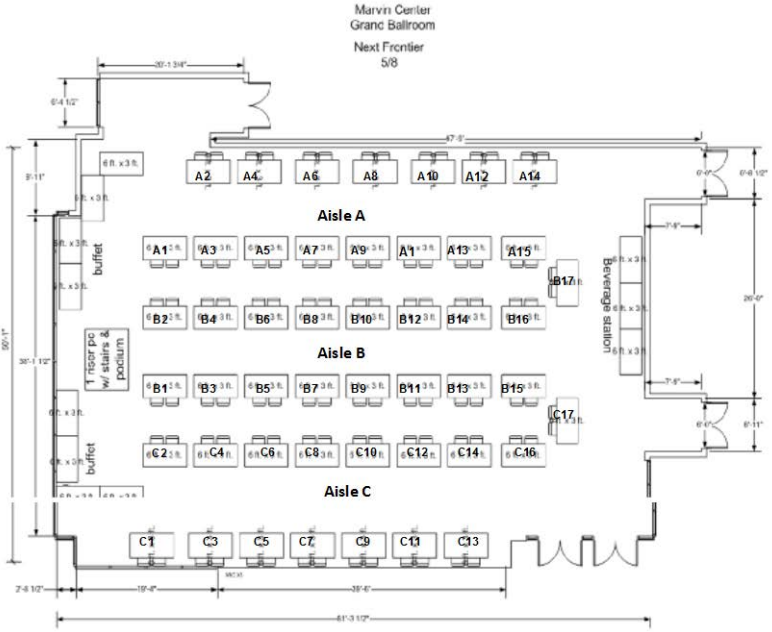

Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
B2
B4
B6
B8
B10
B12
B14
B16
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
C2
C4
C6
C8
C10
C12
C14
C16
Aisle C
C1
C3
C5
C7
C9
C11
C13
buffet
buffet
T riser pc w/ stairs & podium
Beverage station

Networks & Relationships:

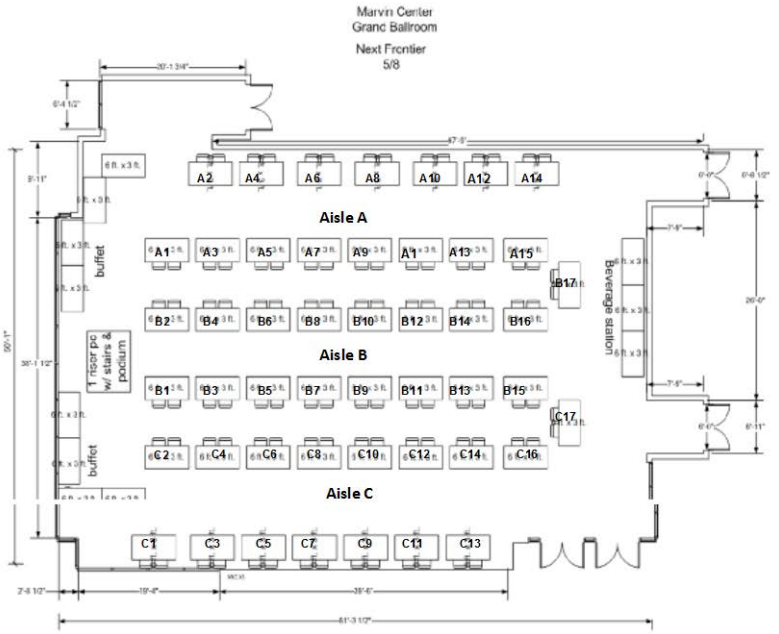
Prices & Costs:


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
buffet
B2
B4
B6
B8
B10
B12
B14
B16
T floor pc
w/ stairs &
podium
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
buffet
C2
C4
C6
C8
C10
C12
C14
C16
Aisle C
C1
C3
C5
C7
C9
C11
C13
beverage station

Regional industries & economies:


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
buffet
Beverage station
B2
B4
B6
B8
B10
B12
B14
B16
1 riser pc w/ stairs & podium
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
C2
C4
C6
C8
C10
C12
C14
C16
buffet
Aisle C
C1
C3
C5
C7
C9
C11
C13

R & D, innovation, commercialization:


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
buffet
B2
B4
B6
B8
B10
B12
B14
B16
T floor pc
w/ stairs &
podium
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
buffet
C2
C4
C6
C8
C10
C12
C14
C16
Aisle C
C1
C3
C5
C7
C9
C11
C13
beverage station

The Opportunity for Innovative Regional 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. In addition, new statistical methods make possible the creation of microdatabases that allow new ways of studying economic behaviors while, when necessary, fully protecting confidentiality. Further, greater policy emphasis on innovation, entrepreneurship, clean tech, and other such building blocks of a 21st century economy is leading to new data collection efforts in these realms.

A plethora of innovative data sources and tools are emerging from a variety of sources, spanning 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 economic well-being. Improved understanding should lead to better designed, more effective policies and programs for stimulating economic growth.

Regions are an important venue for analyzing human activity. Recognized around the world as providing the foundation for economic growth, regions are organically organized, relationship-laden geographies tied by common economic interests and infrastructures that allow the pursuit of those interests. 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, transactions that do not pay much attention to 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 jurisdiction at a moment sometime in the past are not sufficient for generating the understanding that leads to intelligent public and private economic investments.

Innovative sources of economic data, however, offer analysts new ways of seeing—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 do more traditional products. In doing so, they enable analyses that strive to explain how economies work, which in turn facilitate better informed private and public actions to create jobs, income, and profits.

This briefing e-book serves as a guide to the “Innovative Data Sources for Regional Economic Analysis” conference at George Washington University on May 7-8, 2012. The event is an effort to raise the awareness of policymakers, practitioners, and researchers about innovative data sources useful for regional economic development analysis and policy, to build a community of interest on the use of these new sources, and, ultimately, to advance the availability and reliability of useful regional economic data.

These opening remarks consider the potential for innovative data sources to enhance research, practice, and policy making. Our hope is that it will stimulate useful conversations among data analysts and data providers, public and private, that result in a growing community invested in the new generation of regional economic data.

Filling Knowledge Gaps with New Sources of Data

Motivating this e-book and the conference 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 not available or known 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 is to support federal macroeconomic policy and guide the distribution of federal funds to political jurisdictions with particular economic characteristics, such as high unemployment.

While entrepreneurship and innovation are frequently held out as a policy panacea for all that ails economically, we observe many cases in which regions do not to benefit from public and private investments in such efforts and so fail to meet economic challenge and change. In no small part, their failure is due to a dearth of data and information: analysts lack the data to accurately diagnose their regions' economic problems and policymakers lack the tested principles to act on these 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, ones often 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 evidence.

The places most relevant for regional analysis do not neatly fit within pre-determined boundaries. Areas such as Silicon Valley, Route 128, and the Research Triangle – the archetypes of successful regional economies – have complex geographic shapes, build on the location of prominent institutions and firms, are influenced by older transportation routes and land use patterns, and expand out of seemingly idiosyncratic and serendipitous events. These spatial patterns do not acknowledge political jurisdictions,

but rather follow a different logic, one that motivates firms to locate near others with similar products, markets, and workforce skills.

Not only do regional economies blithely ignore town and county boundaries, they don't respect state boundaries as well. At the same time, the geographic clustering of innovative, creative firms often occurs in small places – multi-tenant buildings, neighborhoods, or adjacent industrial parks. These concentrations are invisible when data are available only for larger political units. Data that suggest a cluster, or no cluster, at the county level may indeed mask several geographically, and often technologically, distinct clusters, each grounded in different social relationships and with unique needs. Often, promising early stage activity may be overlooked.

Locating market actors with various characteristics across real economic space requires access to digital data that allows for flexible, user-determined analysis. 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. Researchers can analyze these microdata within unique self-defined economic boundaries.

The innovative firms that are the driving force in regions are themselves fluid and difficult to classify. As firms struggle to survive they often change their activity, however, there is no time or incentive to update their NAICS industrial classification, which has been the main mechanism for understanding industrial activity. For the purposes of understanding new activity information, detail in patent documents provides an idea of where a company is headed; and new product announcements offer a mechanism to understand where firms are placing their bets in the market. Product announcements 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 or relational attributes. We will never gain an understanding of the emerging technologies that have the greatest promise for building new industries by relying on historical classification schemes.

The time lag involved with many data sources means that our analysis is always retrospective, lagging the current reality that we are trying to understand. These limitations reflect the technologies of a prior time, before current administrative data were readily accessible on the Internet, and before we all had at our fingertips the computing power to manipulate them.

Traditional datasets and industrial classification do not describe the relationship among organizations across space. So, for instance, universities, trade associations, business services, and other quasi-government entities are important to innovative activity in regions. These institutions are often the *glue* that holds a regional economy together and provides the foundation for economic vitality. Innovative data sources allow the identification of networks and social relationships between firms, between firms and institutions, and their connections both inside and outside the region, enabling analysts to determine points of leverage for economic expansion.

Long-standing forms of economic data also do not allow analysts 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 easily leads 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.

Efforts to build and maintain regional economic advantage often involves some combination of bottom-up efforts by regional public and private actors and top-down resources invested by the federal government. Such efforts are far more "hit and miss" than would be the case with access to 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 independent and collective efforts of research organizations, federal decision-makers, philanthropic foundations, entrepreneurs and private investors, and economic and workforce development organizations to support, to demand and create a market for, these innovative data sources, to continually redefine the state-of-the-art, to bring to bear the degree of creativity, risk-taking, and entrepreneurship being asked of regional economic actors. The conference organizers hope that this event serves as a springboard for such action.

The Makings of the Conference

Maryann Feldman offered the initial idea for this conference in her first meeting with Andrew Reamer a year ago. E.J. Reedy of the Ewing Marion Kauffman Foundation recognized their mutual and synergistic interests and introduced the two and encouraged their collaboration and development of a joint proposal. Andrew and Maryann sketched out the framework for a workshop in that proposal, which the Kauffman Foundation funded.

Maryann Feldman has long been interested in the use of nontraditional data sources in regional economic research. Her early work relied on a unique, federally-funded dataset of new product introductions in 1982. This experimentation continued: Feldman and Gil Avnimelech used data from LinkedIn to track parent-founder relationships and Feldman and Zoller used S&P Capital IQ to describe network relationship patterns that lead to new business development.

The relevance of new data sources is demonstrated Maryann Feldman and Nichola Lowe's work with a relational database created at the University of North Carolina in 1991 to follow the progress of entrepreneurial firms in the region surrounding the Research Triangle Park (RTP). William A. Little, then Acting Provost, long-time faculty member in the Chemistry Department, and original board member of RTP, set out to refute a claim that the region had few new start-ups. Because of Bill Little's prominence in the region, he was in a good position to be able to track companies and continued to collect data until his death in 2009, at which time the database had grown to 1,800 firms.

Bill Little's effort was picked up by Maryann Feldman and Nichola Lowe, who first vetted the data against various external sources to ensure accuracy and then augmented firm records with data from other sources. Their academic objective was to study in depth the dynamics of this one economically important region, arguably a place created through public efforts. The larger outcome of the project is a replicable methodology for creating this type of micro-level database and an appreciation of the promise of new digital data sources.

Andrew Reamer has long appreciated the importance of data to inform policy. Between 1984 and 2004, Andrew ran two regional economic development consulting firms, the work of which was based on gathering and integrating data from multiple public and private sources (from big books and mimeographs in the beginning) to tell a meaningful story about a region's economic performance and competitiveness, the challenges ahead, and how to address them. In the late 1990s, his work began to focus more on the extent to which federal statistical agencies are meeting the needs of regional data users and how they might better meet those needs. At GW, a good part of his efforts involve encouraging sufficient congressional funding of federal statistical agencies and those agencies' development of data products that users value.

Andrew and Maryann proposed a workshop with representatives of 15-20 data sources and 50-60 data users. As they began work last winter, they quickly realized that relevant innovative data sources were more numerous and diverse than originally expected. This diversity was both by subject and type. Regarding the latter, they found innovative data sources that aim 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 are many examples at the conference, including

Amazon Web Services, Moody's Analytics, Windows Azure Marketplace, Factual, TheDataWeb, and Data.gov)

The organizers also thought that good opportunities existed for knowledge exchange, collaboration, and contractual relations among these various types of providers. Consequently, they decided to significantly expand the number of presentations.

Regarding the invitation of data users, the organizers thought that a by-invitation-only approach would reduce the conference's value because participants would be limited to users known to the organizers and likely to each other. Conversely, it seemed that an open invitation, to bring together people who do not know each other and whom the organizers do not know, would better stimulate new relationships among users and providers that would lead to 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 have open registration and relabel the workshop as a conference. The response from data stakeholders has been substantial and the conference will be 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 has allowed the event to take place in its current form.

The Briefing Book and Other Conference Publications

This briefing book's primary aim is to assist conference attendees in deciding which innovative data sources to visit and to serve as a reference after the event (though one with a short shelf life in light of the rapid pace of change). Aligned in alphabetical order, each

data source's 2-3 page overview follows a somewhat similar structure of subject headings. Each data source is categorized by topic or type, with a listing by category in the appendix.

After the conference, user blogs and data source evaluations will be compiled into a published set of proceedings. The organizers also will prepare an assessment of the potential utility of innovative data sources for EDA's program evaluation process. Currently, Maryann is leading a project to provide EDA with recommendations for redesigning its evaluation process. Andrew is GW's lead as the project subcontractor.

Conclusion

The availability of economic data and the tools to analyze them has never been greater. At the start of Maryann and Andrew'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 analysis and insights previously beyond anyone's imagination and offer the opportunity for greater and more sophisticated understanding that will be important for policy and research.

Ultimately, the conference's aim is to create the conditions for learning, connection, and serendipity. We hope that participants believe that this aim is fulfilled.

Altmetrics Total-Impact

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

Overview: Total-Impact measures the impacts of scholarly research through gathering data on 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 Regional Analysis: Compare impacts of departments and institutions within a region and across regions, by research field or topic

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Data providers can use AWS to host, compute, and analyze Big Data sets. 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)

Provide feedback for this exhibitor? Click here.

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, 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 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 with telephone and in person interviews

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

Potential Uses for Regional Analysis: Determination of workforce, job, and journey-to-work characteristics, with substantial disaggregation by age, gender, race, and ethnicity**.**

For Additional Information:

- Website: American Community Survey

Provide feedback for this exhibitor? Click here.

Business Dynamics Statistics

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

Overview: Business Dynamics Statistics (BDS) provides for measures of net and gross job flows associated with entering, exiting, expanding, and contracting establishments including measures of job creation and destruction. 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 Regional Analysis: Ascertaining 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

Provide feedback for this exhibitor? Click here.

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, for the nation and states. 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 Regional Analysis:

- Identify patterns of gross job creation and destruction by industry sector
- 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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: 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.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 key-word 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 Regional 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

Provide feedback for this exhibitor? Click here.

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 and 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 funded by angel investors, government grants and incubators to private equity and venture capital backed firms to 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
  - Angle
  - 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 & company websites and tens of thousands of other sources identifying investment and M&A news related to private companies of interest.

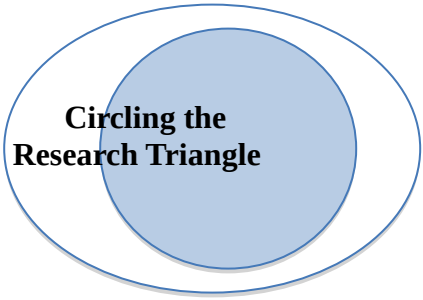
Access: Subscription required for current data**;** Historical VC and angel reports are available for free

Potential Uses for Regional 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 & 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 & 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)

Provide feedback for this exhibitor? Click here.



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, 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 3,900 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

Uses for Regional 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 places.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 for 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: 358 index values; 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 within a state.

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 Regional Analysis: 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

Provide feedback for this exhibitor? Click here.

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. The openness derived from Data.gov will 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 Regional 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 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

Provide feedback for this exhibitor? Click here.

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

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Enhances accessibility and usability of relevant datasets; facilitates integration 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)

Provide feedback for this exhibitor? Click here.

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, 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 Regional Analysis: Users are able to create scientific decision support systems.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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, demographics — as well as data on occupational skills, education, training, and for a specific region or industry.

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 source 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 over from approximately 80 federal, state, and private 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 Regional Analysis: The data are used to research and understand employment trends, education and economic development, and dynamics.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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, 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. However, the origin-of-movement is not always the place where the good was produced, especially when the exported goods get 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 NAFSA: 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 Regional Analysis: 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)

Provide feedback for this exhibitor? Click here.

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: Facutal 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 Regional Analysis: Customized 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)

Provide feedback for this exhibitor? Click here.

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

For Additional Information:

- Website:
  - GeoIQ
  - GeoCommons 2.0

Provide feedback for this exhibitor? Click here.

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, MSAs, 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 Regional Analysis: Identify scope of green goods and services activity at the state level

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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

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 Regional 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)

Provide feedback for this exhibitor? Click here.

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 to be 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 to be released in August 2012 will provide 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

Lawyers

- Name
- Addresses of law firms

Underwriters

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

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

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 Regional Analysis: Examination of 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)

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Determination of 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)

Provide feedback for this exhibitor? Click here.

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 including 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 Regional Analysis: Track evolution of region's establishments by location or industry; identify contributions of young vs. old firms or small vs. large firms

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 frequency 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 Regional Analysis: QWI analyzes demographics and wages of newly hired versus other workers in the same industry and employment trends at sub-state geographies.

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

Provide feedback for this exhibitor? Click here.

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; further enhancing 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 & 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 Regional Analysis: Mendeley may be used as both a data source to track regional research activity and as a collaborative platform to enhance regional collaboration.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Monsters.com or from resume databases loaded into the platform for analysis

Access: Hosted cloud solution

Potential Uses for Regional 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)

Provide feedback for this exhibitor? Click here.

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 years

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 Regional Analysis: Analyze multiple dimensions of regional economies. Forecasting.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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
- Type of establishment
- Employment
- Estimated annual sales (firm-level)
- Dun & Bradstreet Credit Score
- Dun & Bradstreet PayDex scores
- Special indicators

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 Regional 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)

Provide feedback for this exhibitor? Click here.

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 and undergraduate degree to 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 Regional Analysis: 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

Provide feedback for this exhibitor? Click here.

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 lined 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 expert 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 Regional Analysis: Assess regional occupational and skills clusters, using crosswalk between occupations and skills; identify workforce KSAa and GWAs; 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)

Provide feedback for this exhibitor? Click here.

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 & 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 Regional Analysis: The methodology can be applied at the state or MSA level to monitor price trends and provide real-time information on inflation.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Recent research topics include 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

Provide feedback for this exhibitor? Click here.

U.S. Department of Commerce
Bureau of Economic Analysis

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, and other goods and services as well as for 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 Regional Analysis: Compare price levels across different geographic areas in general and for specific groups of goods and services. Adjust regional measures of income and output for price level differences.

For Additional Information:

- "Regional Price Parities by Expenditure Class for 2005-2009 in th 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

Provide feedback for this exhibitor? Click here.

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 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 and 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 Regional Analysis: Examine associations of innovative behavior with establishment survival and employment growth; analyze the role of location (urban verses rural), firm size, local institutions, and proximity in innovation activities and innovation networks

For Additional Information:

- Website: Rural Establishment Innovation Survey

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Assessment of trends in science and technology over time and across states

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Create regional corporate databases, by industry, with detailed financials. Identify linkages and activities among companies, investors, and subsidiaries in specific regions.

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

Provide feedback for this exhibitor? Click here.

Science of Science (Sci^2) Tool

Categories: Data analysis and visualization tools

Overview: The Science of Science (Sci^2) 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 Sci^2 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 Sci^2 tool team welcomes developers to contribute their own extensions to the Sci^2 project, see algorithm developer guide.

Access: Open source, Apache 2.0 license

Potential Uses for Regional Analysis: Visualizations of any form of data on individuals and places.

For Additional Information:

- Website: Sci^2
- Contact: Robert Light (lightr@indiana.edu)

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Researchers can use the data to analyze expertise, education, networks, and research by an area or institution.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Assess strengths, capabilities and research topics of researchers and research institutions. Identify researchers with relevant expertise and interests. Examine research awards and explore funding opportunities. Enables identification of collaborative partners.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

ELSEVIER

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 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 Regional Analysis: STAR METRICS enables estimation of the multiplier effects of R&D spending: direct impacts of R&D employment and spending on local exhibitors,

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)

Provide feedback for this exhibitor? Click here.

NATIONAL CENTER FOR EDUCATION STATISTICS

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 made 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 Regional Analysis: SLDSs will allow analysts to trace the workforce outcomes (in terms of employment, industry, and wages) of students' education. Comparisons possible 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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Assess strengths and weaknesses of regional economies, particularly with regard to brainpower and innovation

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 collected by the Association of University Technology Managers (AUTM) from participating academic institutions.

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 speadsheets 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.

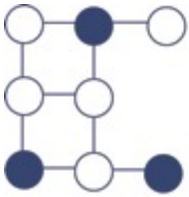
Access: Subscription

Potential Uses for Regional Analysis: 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

Provide feedback for this exhibitor? Click here.



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
- Opportunities for promotion & influence

Firm Performance

- Revenues
- Exports
- Employment
- Valuation

- Research & business linkages
- Technology services
- Complementary business services
- 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 Regional Analysis: Measurement of impacts of investments in research, innovation, and business support by funder, class of funder, and region. 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
  barge@theevidencenetwork.com

Provide feedback for this exhibitor? Click here.

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. Subtantial information available on individual firms as well.

Timeframe: Annoucements 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 Regional Analysis: Identify extent and nature of product introduction trends by region

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)

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Detailed identification of a region's economic structure. Systematic comparison across regions. As cluster category definitions are standardized, they are most useful when looking across all regions.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 subawards

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 that allows user-defined downloads, 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 Regional Analysis: 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

Provide feedback for this exhibitor? Click here.

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. It 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 Regional Analysis: Regional analysis of patent and trademark activity.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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

- Material transfer agreements
- Consortia agreements
- Sponsored research by industry
- Clinical trials
- Service to external clients

Developing the regional and national workforce

- Student employment on funded projects
- Student economic engagement
- Student entrepreneurship
- Alumni in the workforce

Knowledge incubation & acceleration programs

- Incubation & accerlation 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 Regional Analysis: 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

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

Wanted Analytics

Category: Labor markets

Overview: WANTED *Analytics*™ provides real-time intelligence on labor markets. Clients use WANTED *Analytics*™ 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, education 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 Regional Analysis: Analysis of regional labor market characteristics and dynamics.

For Additional Information:

- Website: Wanted Analytics

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

Provide feedback for this exhibitor? Click here.

Web of Knowledge

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

Overview: *Thomson Reuters Web of Knowledge℠* 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 Knowledge℠* and *Web of Science℠* 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 Science℠
- 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 Regional Analysis: Identification of the impacts of regional scientific research activity on the works of others. Identification of regional and interregional networks and relationships.

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Access to a large number of datasets relevant to regional analysis

For Additional Information:

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

Provide feedback for this exhibitor? Click here.

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 Regional Analysis: Identify dynamics of regional economic change, such as expansions of existing establishments, impacts of relocating firms, 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

Provide feedback for this exhibitor? Click here.

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
Sci^2: 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
STATT: 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.

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

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.

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 geogpraphy 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

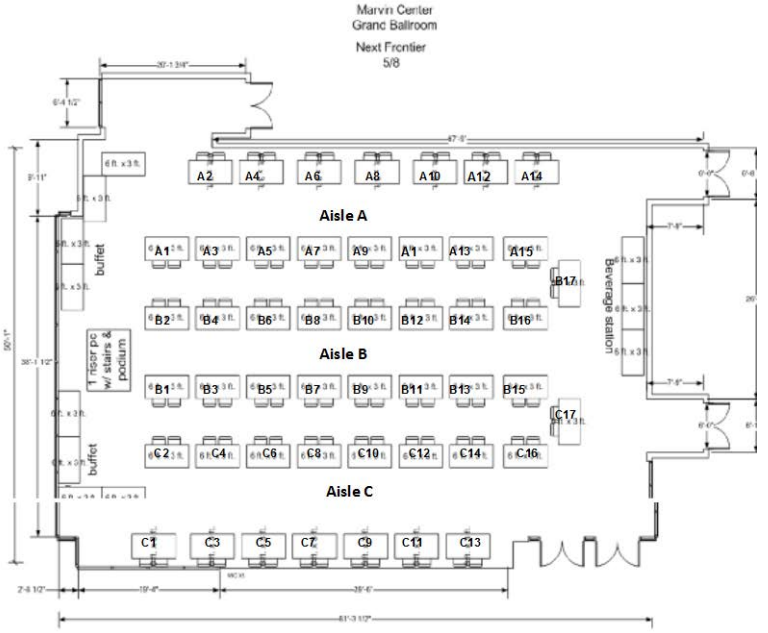
Time Series Data: Data on an event at regular time intervals, analogous to longitudinal or panel data

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.

Innovative Data Source, Alphabetic Order


Marvin Center
Grand Ballroom
Next Frontier
5/8
A2
A4
A6
A8
A10
A12
A14
Aisle A
A1
A3
A5
A7
A9
A11
A13
A15
B17
B2
B4
B6
B8
B10
B12
B14
B16
Aisle B
B1
B3
B5
B7
B9
B11
B13
B15
C17
C2
C4
C6
C8
C10
C12
C14
C16
Aisle C
C1
C3
C5
C7
C9
C11
C13
buffet
buffet
1 riser pc w/ stairs & podium
Beverage station

Innovative Data Source, Table Order

Burning Glass: Labor/Insight	A01
American Community Survey 5-year Estimates	A10
RDC: Research Data Center Program	A11
BDS: Business Dynamics Statistics	A12
GGS: Green Goods and Services Survey	A13
PROGRIS: Globalization & Regional Innovation	A14
Brookings Sub-National Export Series	A15
EMSI's Products	A02
Monster Government Solution's	A03
Geo-IQ Products	A04
NSC Tools	A05
Sci2 Tool	A06
DataWeb for the DataFerrett	A08
STATS America	A09
USPTO Research Datasets	B01
COLA for the New Supplemental Poverty Index	B10
SciVal	B11
NETS Database	B12
SciENCV: Sicence Experts Networtk and CV	B13
S&P Capital IQ's Products	B14
S&E Indicators State Data Tool	B15
CB Insight's Products	B16
BED: Business Employment Dynamics	B17
LBD & SnyLBD: Longitudinal Business Database	B02
USAspending.gov	B03
Wanted Analytics	B04
ThomasNet’s Product News Room	B05
O*NET Data Collection Program	B06
The Evidence Network	B07
Regional Price Parities	B08
Price Stats Products	B09