Census Bureau Innovation Measurement Initiative

GUIRR Webinar
5/25/2016

Ron Jarmin

U.S. Census Bureau

Broader Efforts at Modernizing Economic Measurement

- Users demand more timely and granular economic statistics
- Hard to meet these needs with surveys alone
- New “Big Data” sources hold the promise of delivering economic and social measurements with greater detail and improved latency.
- Census and other statistical agencies exploring ways to modernize products and operations using these rich new data sources and new analytic tools

Census Bureau activities

- Established Center for Big Data Research and Applications
- Exploring
  - alternative source data for retail trade estimates
  - passive data collection for business surveys
- Partnered with academia to develop Big Data class
  - Textbook forthcoming edited by Foster, Ghani, Jarmin, Kreuter and Lane
- Census – MIT Big Data Workshops
- NCRN and other research activities
- ***Innovation Measurement Initiative (IMI)***

IMI Background

- Census Goals:
  - Improve measurement of small but important sector of the economy
    - Understand “productivity” of science investments
    - Understand impact on the broader economy and its role in innovation, entrepreneurship and job growth
  - Learn how to collaborate with data providers to deliver data products they value
  - Prototype project that can be scaled and extended to other sectors of the economy

IMI Background

- Innovative Aspects:
  - Collaboration with the University of Michigan's Institute on Research in Innovation and Science (IRIS)
  - Experiment with utilizing "fat pipe" of data for a sector of the economy
    - The University data is complementary to business and household data at Census

University Administrative Data: IRIS-UMETRICS

- Institute for Research on Innovation & Science (IRIS) UMETRICS collects administrative data detailing expenditures on funded research projects at universities.
- Includes transaction level data on
  - **People** charging to grants (e.g., faculty, staff postdocs, grad students, undergrads)
  - Purchases of **equipment, supplies and services**

IRIS Operational structure

Universities contribute data.

support infrastructure

campus-specific and ag

products

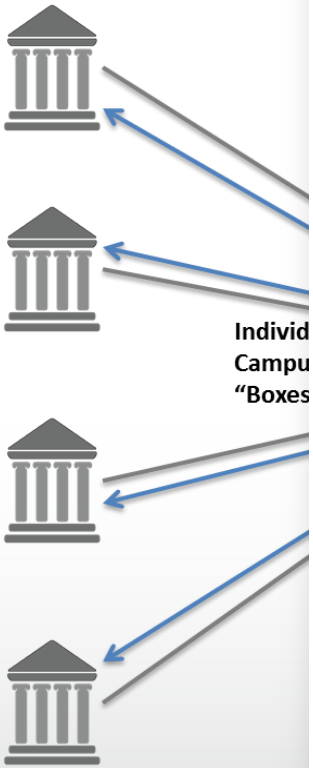
Approved nodes

Approved researchers

y access de-identified,

ate datasets

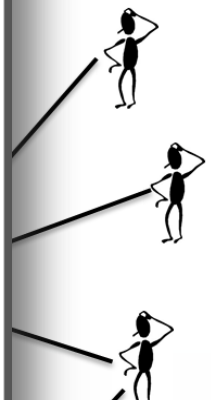
Create New research and reports

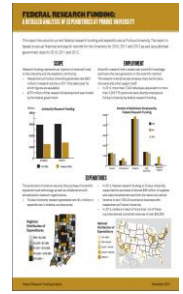


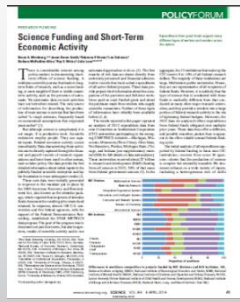
Individ

Campu

"Boxes







"NPR membership"

have accessed data

ASSOCIATION OF AMERICAN UNIVERSITIES

President

July 23, 2014

Professor Jason Owen-Smith
Barger Leadership Institute Professor of Organizational Studies & Sociology
University of Michigan
426 Thompson St. # 3306,
Ann Arbor, MI 48104

Dear Professor Owen-Smith:

As president of the Association of American Universities (AAU), I write in strong support of the Institute for Research on Innovation and Science (IRIS). AAU is an association of 60 U.S. and two Canadian leading public and private research universities. AAU's member institutions perform nearly 60% of federally-funded university research and grant nearly half the nation's doctoral degrees. As universities respond to increasing pressures, the Institute has the potential to be immensely helpful in at least three ways, by:

1. Allowing universities to more comprehensively document the outcomes and impacts of research. This includes not just traditional measures like publications and patents but also economic impacts (e.g., through vendor purchases or technology transfer) and the effects of training individuals. Such approaches may include documenting impacts for units smaller than an institution (e.g., a school or department), as well for sets of institutions (e.g., within a region).
2. Assisting institutions in better understanding their own research portfolios so as to maximize opportunities for growth and collaboration while also allowing efficient use of resources.
3. Providing general insights into the ways science progresses and innovation occurs. Such studies will be facilitated by the ability to follow people and their collaborations across institutions, groups, and sectors over time. The resulting insights, which will be underpinned by high-quality research, can inform not just institutions themselves, but also policymakers, funders, and society at large.

One concern institutional administrators have expressed in the past is that, as more data becomes available, universities may be unfairly benchmarked or subjected to accountability measures that are not commensurate with institutional missions. This project, by setting up an independent research institute to store data securely and by utilizing top scholars to analyze and report objectively on these data, takes major steps to assuage this concern.

Several AAU member institutions, under the leadership of the Committee on Institutional Cooperation (CIC), have already provided data to early stages of this project. More than half of AAU's member institutions are registered to provide nearly identical data through the STAR METRICS Initiative. These institutions have committed time and resources, reflecting their inherent understanding of the value of these data. AAU will provide information on the potential of the Institute, facilitate conversations between institutional leaders, and encourage our member institutions to participate through contributing data and infrastructure support.

www.aau.edu • 1200 New York Avenue, NW • Suite 550 • Washington, DC 20005 • 202.408.7500 • 202.408.8184 fax

UMETRICS @ IRIS

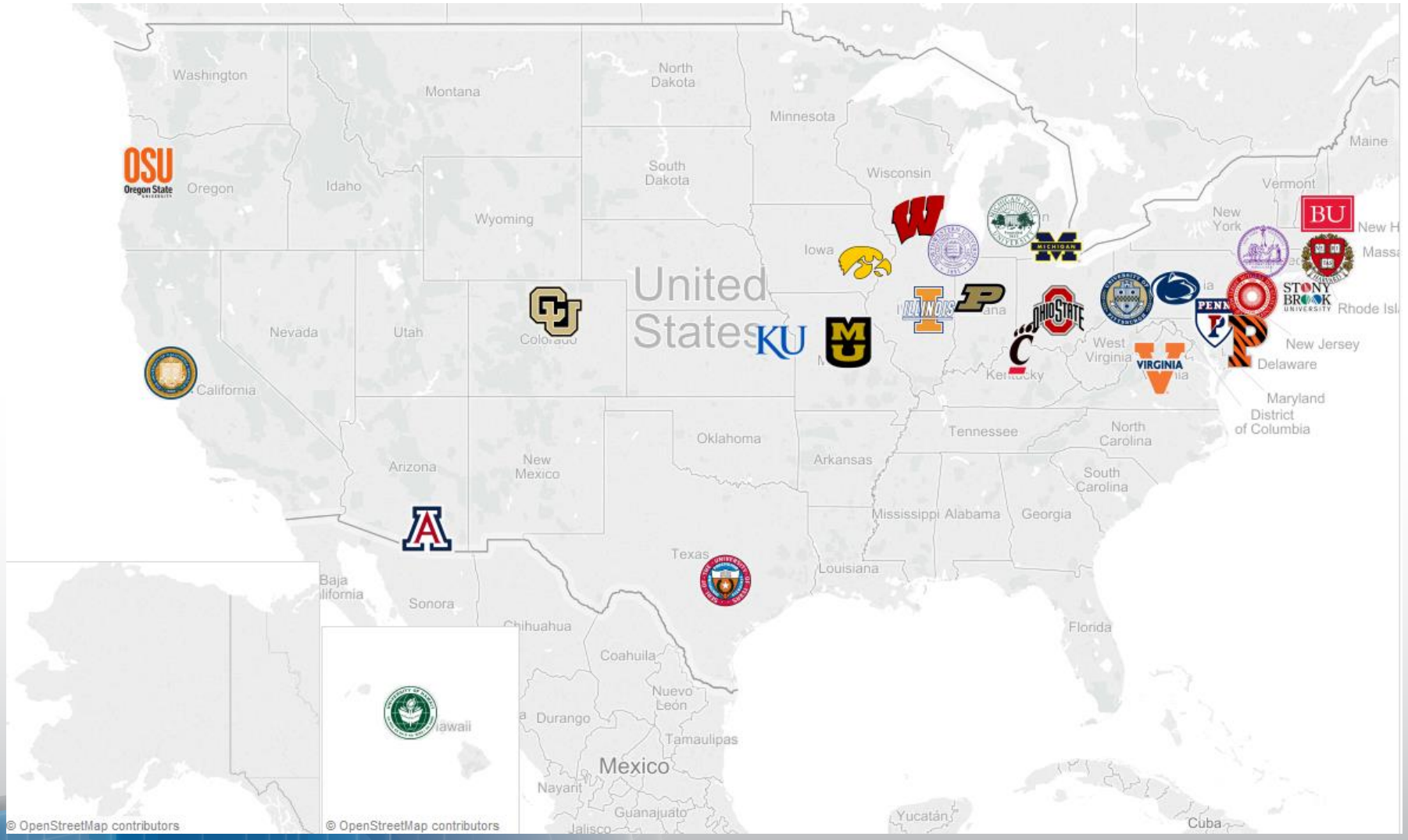
Goal: National Coverage in 3-5 Years

- >150 institutions
- All 50 states
- >90% R&D Spending

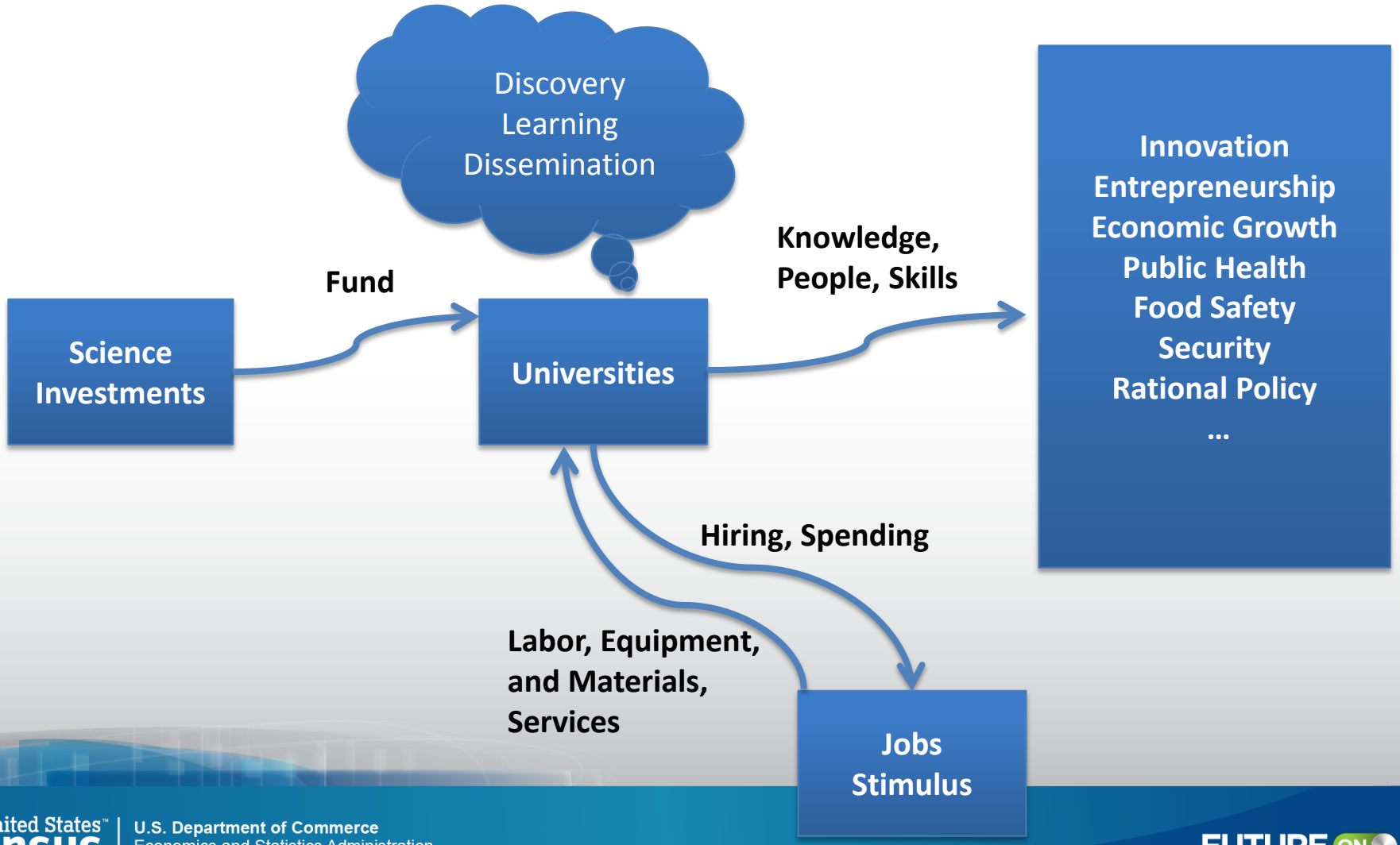
Seed Funding for IRIS infrastructure:

Ewing Marion KAUFFMAN Foundation

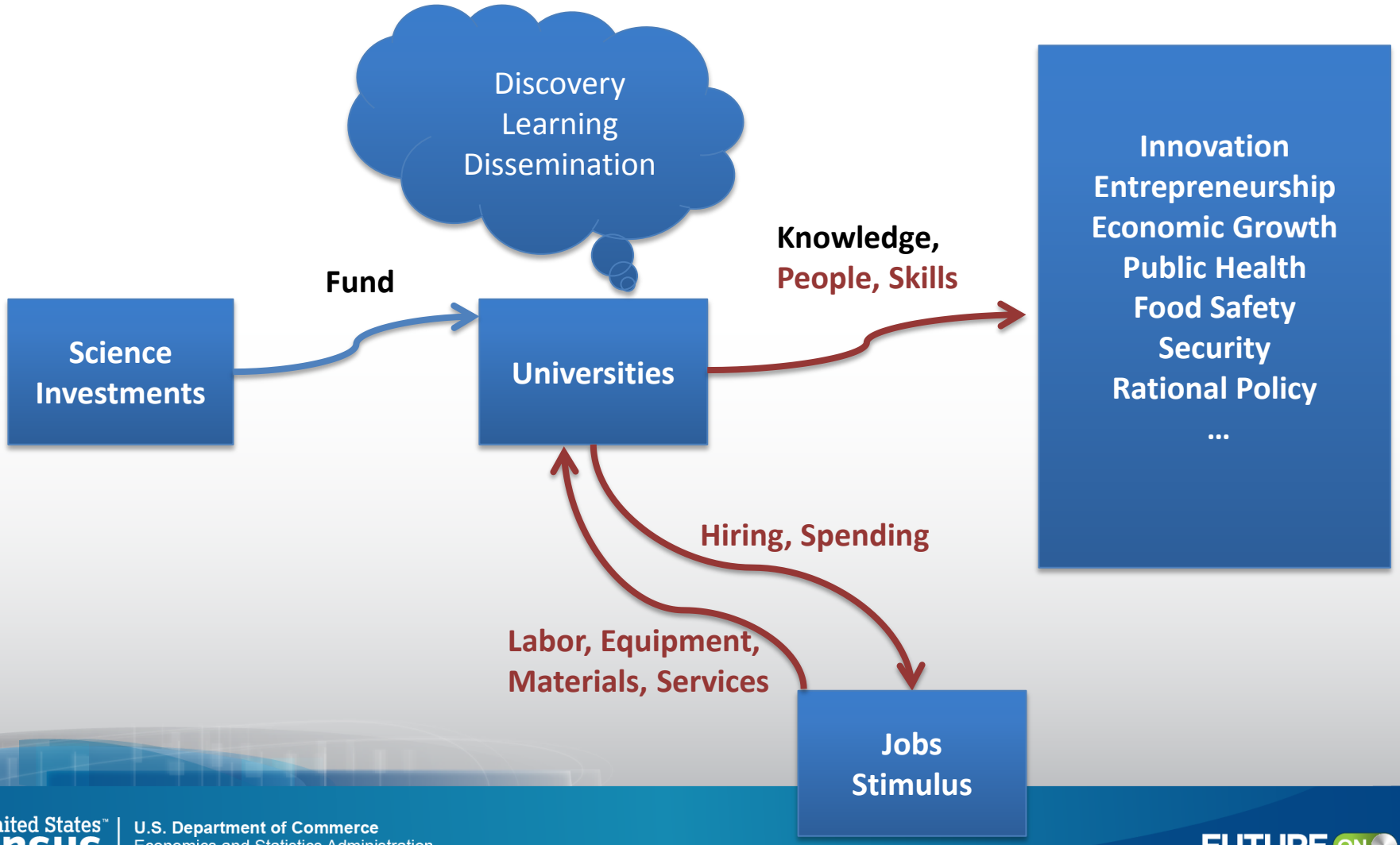
IRIS Participants



Framework

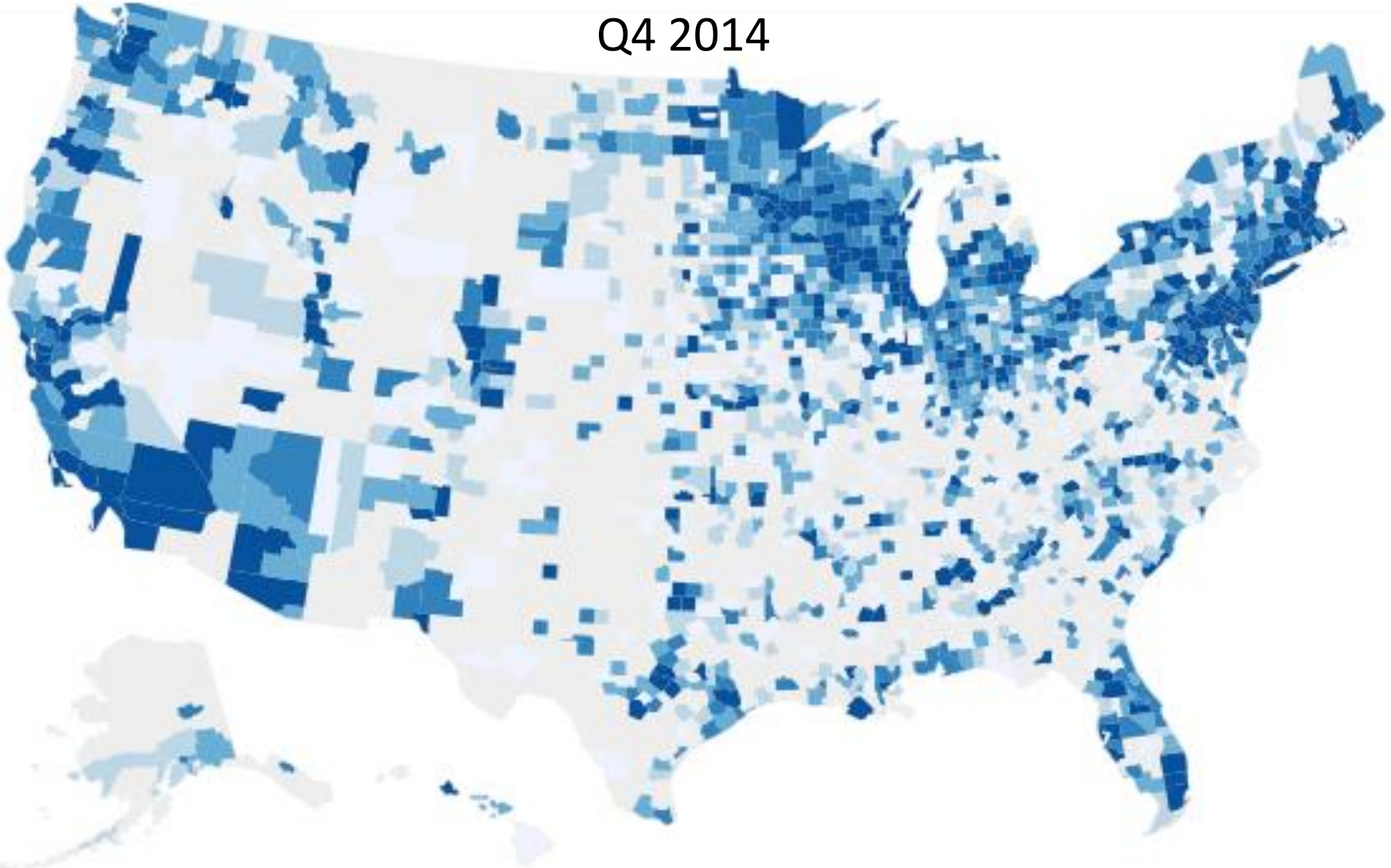

Discovery
Learning
Dissemination
Fund
Science Investments
Universities
Knowledge, People, Skills
Innovation
Entrepreneurship
Economic Growth
Public Health
Food Safety
Security
Rational Policy
…
Hiring, Spending
Labor, Equipment, and Materials, Services
Jobs Stimulus

Framework


Discovery
Learning
Dissemination
Fund
Science Investments
Universities
Knowledge, People, Skills
Innovation
Entrepreneurship
Economic Growth
Public Health
Food Safety
Security
Rational Policy
…
Hiring, Spending
Labor, Equipment, Materials, Services
Jobs Stimulus

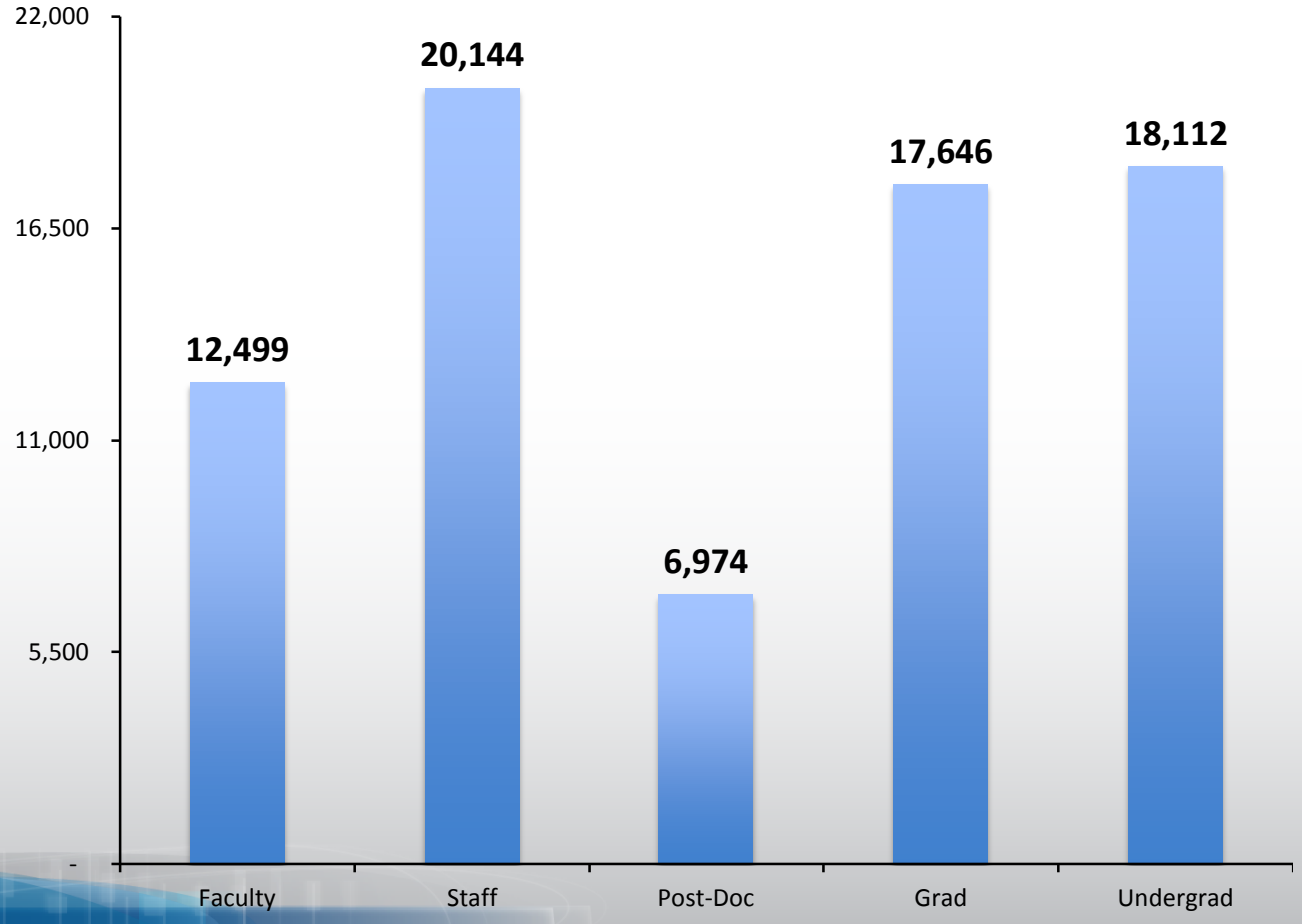
Data on Spending

$1.949 Billion in Direct Cost Vendor Purchases from 9 CIC Universities, Q3 2012-Q4 2014



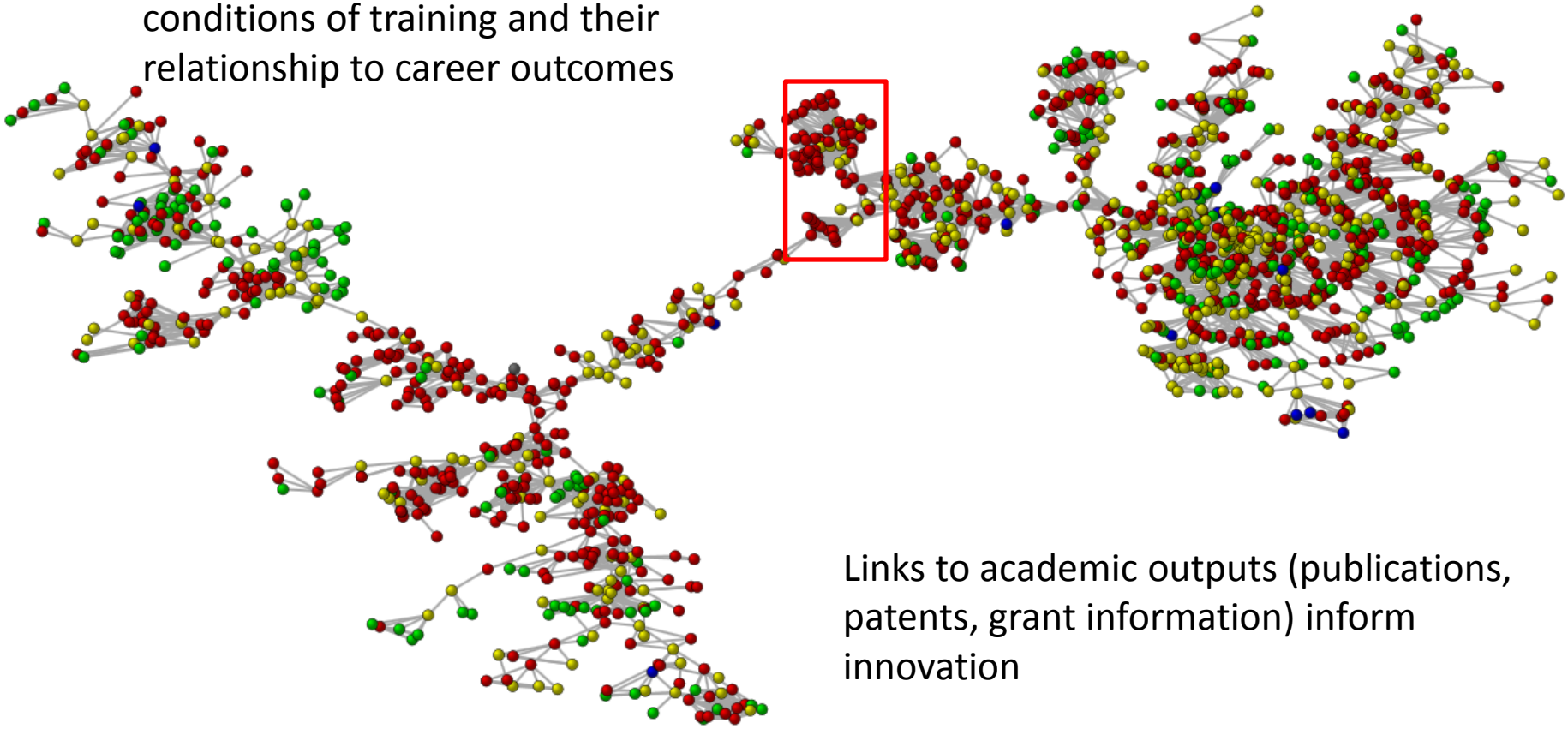
Data on People

75,375 employees by type, 9 CIC Universities, Q3 2012-Q4 2014



Data on Networks

Networks provide insights into conditions of training and their relationship to career outcomes

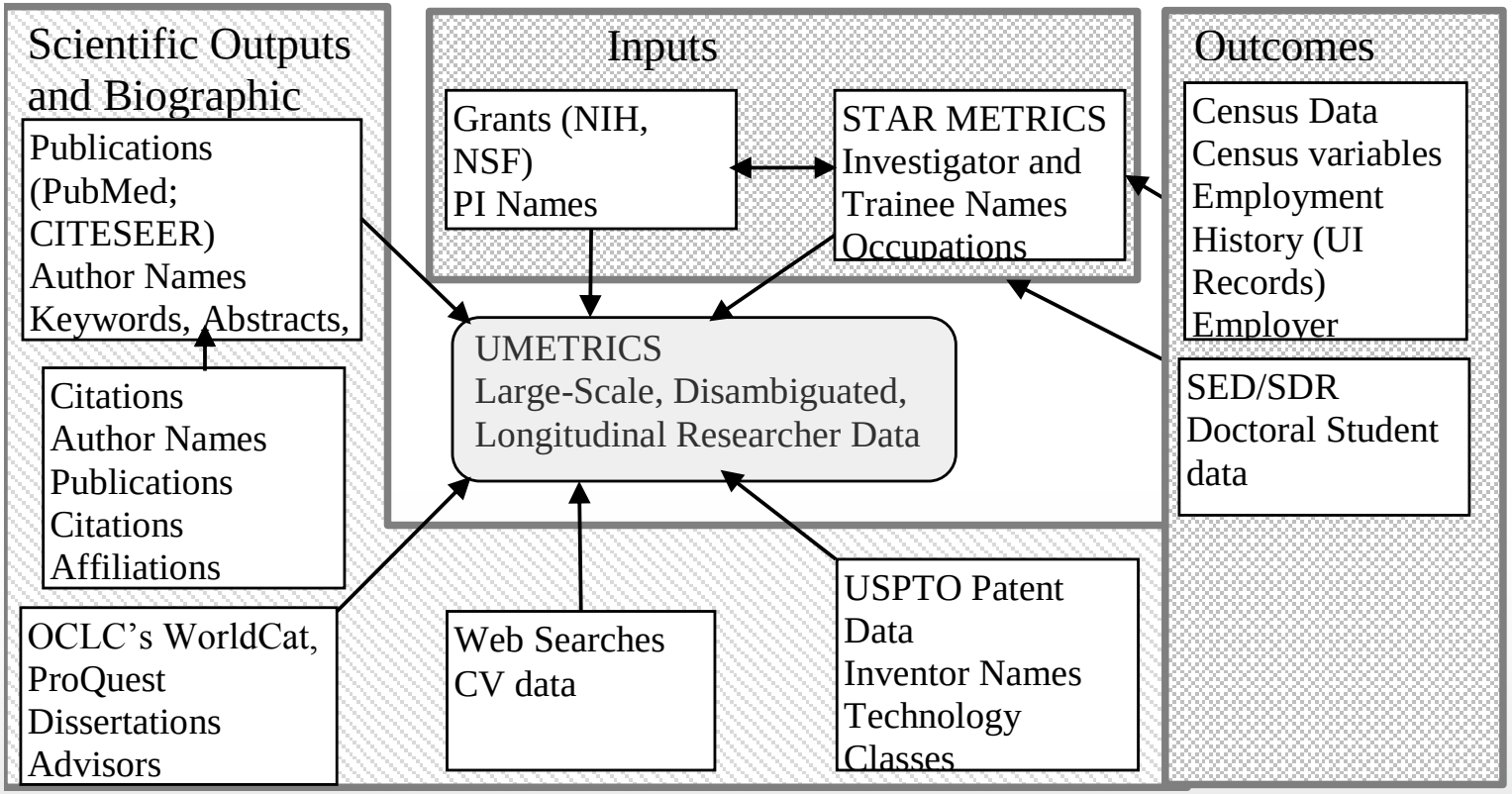


Links to academic outputs (publications, patents, grant information) inform innovation

Products

- Hot Reports tailored for the universities
- Other public use products TBD
- Research Papers
- Restricted Use Data made available through the Federal Statistical Research Data Center network

The Emerging Large-Scale, Disambiguated, Longitudinal Researcher database

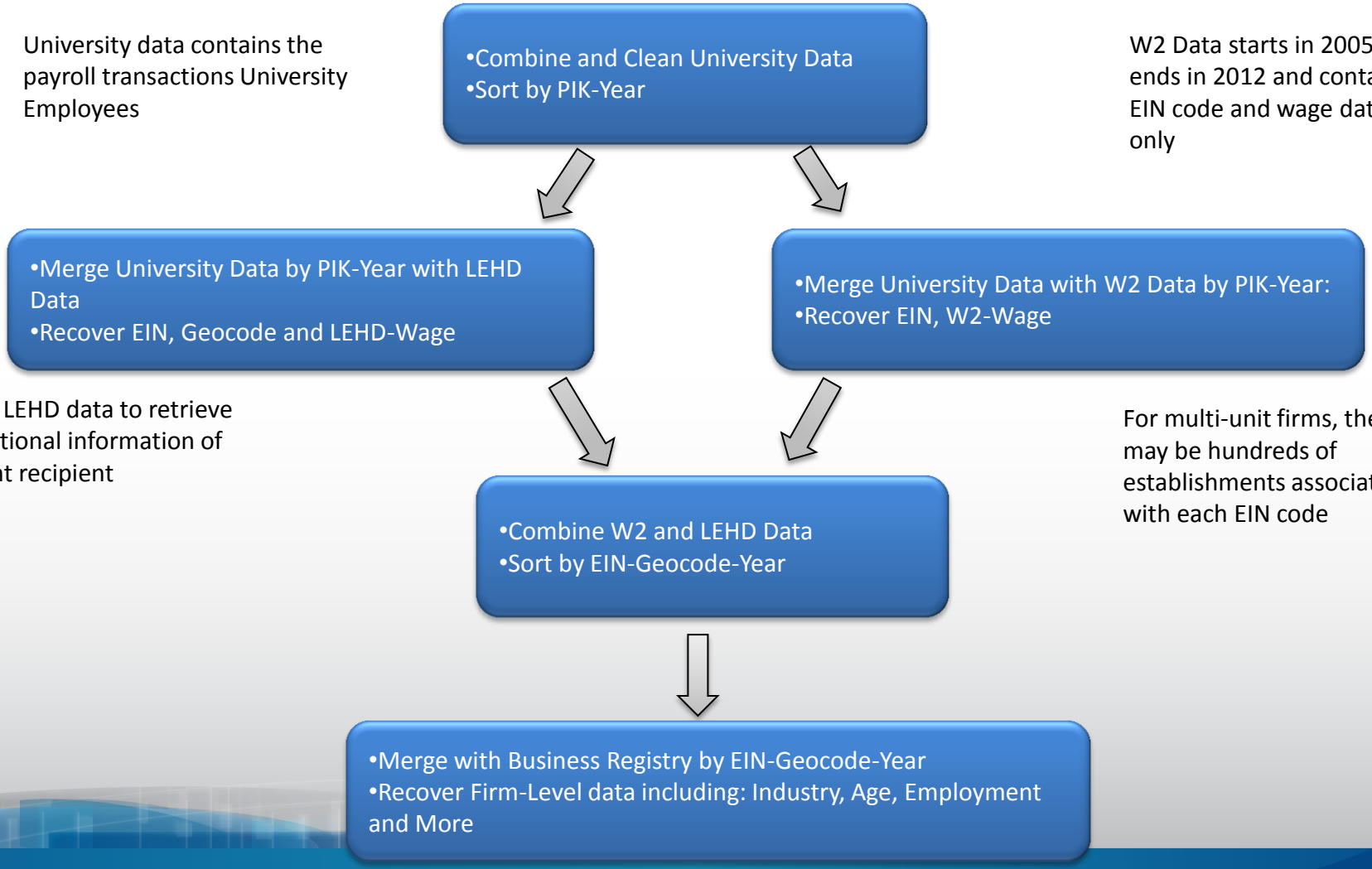


Key Census Data Assets

- Business Register (BR)
  - Universe of U.S. non-agricultural businesses and the source of data from which all other economic data are ultimately created
  - Key data provided: Industry Classification (NAICS), Geographic data, Employment, Payroll, EIN Codes, Available from 2002-2012
- Longitudinal Business Database (LBD)
  - Universe of employer businesses, unique establishments, the LBD covers all industries and all U.S. States linked over time
  - Key data provided: Industry Classification (NAICS), Geographic data, Employment, Payroll, Firm Age, Available from 2002-2012
- Integrated Longitudinal Business Database (iLBD)
  - Universe of non-employer businesses with links to employer universe
  - iLBD records are identified by either PIKs or EINS, 85-88% are PIKs and 12-15% are EINS
  - Key data provided: Industry Classification, Gross Receipts, Geographic data, Available from 2002-2010
- Longitudinal Employer-Household Dynamics (LEHD)
  - Employee-Employer linked dataset
  - Key data provided: EIN-Geocode Linkage, Wage Data, Available from 2002-2010
- W2 Data
  - Key data provided: PIK, Wage Data, Available from 2005-2012

Matching Process – Employees

University data contains the payroll transactions University Employees



W2 Data starts in 2005 and ends in 2012 and contains EIN code and wage data only

Use LEHD data to retrieve locational information of grant recipient

For multi-unit firms, there may be hundreds of establishments associated with each EIN code

Matching Process – Vendors

- Fuzzy Name and Address Matching
  - Data cleaning, standardizing, and creating match-codes
  - Run cleaned files through SAS block-matching programs (utilizing SAS Data Quality package)
- Match process proceeds in two steps:
  - Vendors are first matched to the BR
  - Best BR match(es) by year are then matched to the LBD
  - Analysis using employment weighted probabilistic selection for ambiguous matches
- Final matched dataset is a vendor-year panel
  - Having street address improves accuracy of establishment identification
  - Vendor match rates range from 40% to 70%, dollar weighted average 60%
  - Future improvements – cleaner vendor identifiers, focusing on certain types of purchases

Employee Outcomes

Matching

Year	Total		Total		Total		Total
2010	13,068	→ PIK-ize	10,126	→ Matched to LBD (left university)	6,097	→ Doctoral Recipients	578
2011	19,323		14,658		9,335		692
2012	22,478		17,010		10,723		713
Sum	54,869		41,794		26,155		1,983

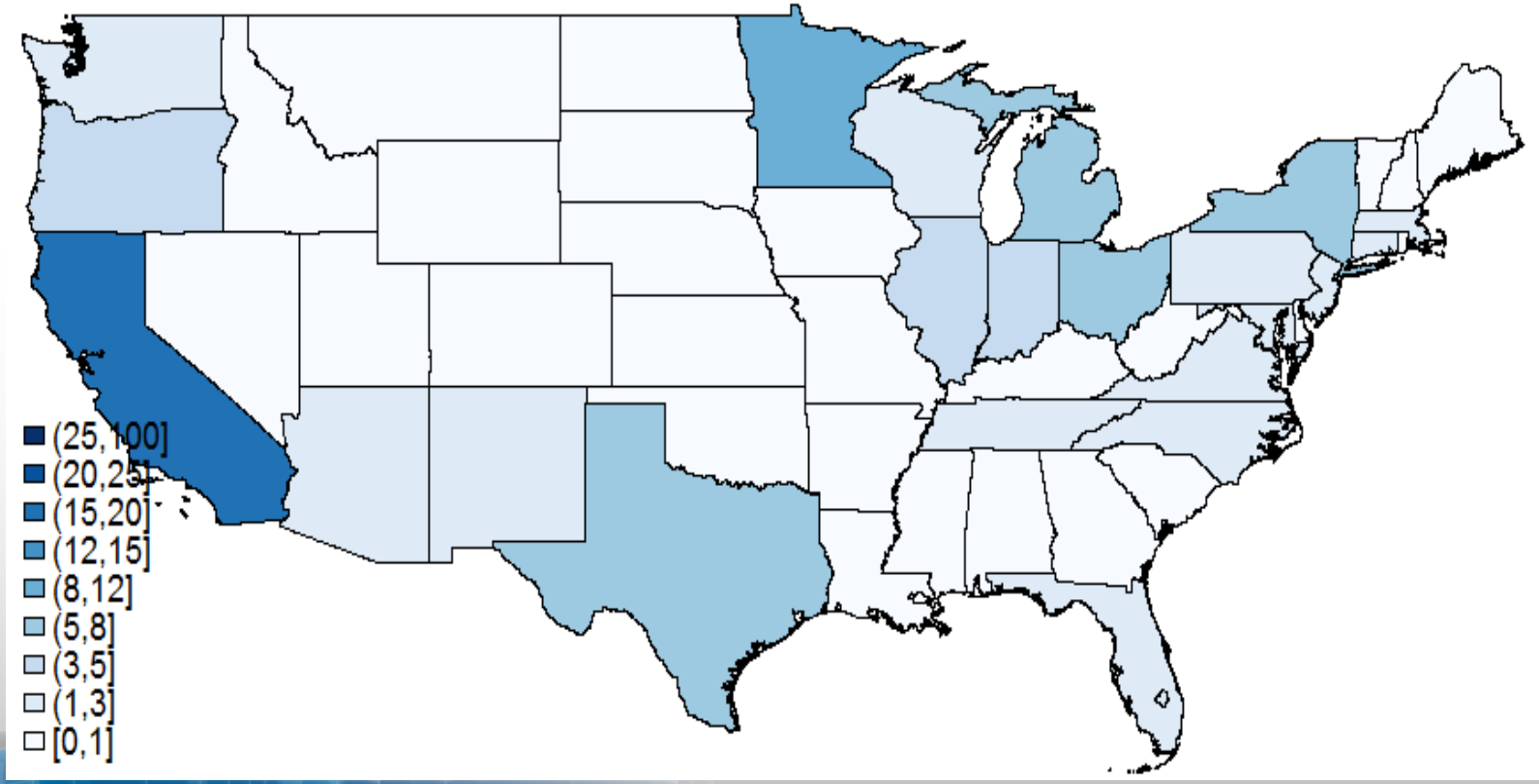
Job Placements - 1 Year After Leaving Institution

(Zolas et. al. *Science*, Dec. 2015)

Table 1. Postgraduation employment of UMETRICS doctoral recipients who were paid by research grants and left the university between 2009 and 2011. The national workforce distribution is calculated from all employment in all establishments covered by the Census's LBD between 2010 and 2012.

	Doctoral recipients placed in sector (%)				
	Industry		Academia	Government	All
Locale and small	R&D firms	Non-R&D firms			
Placed within sector	17.0	21.7	57.1	4.1	100.0
National sample (*M*)	10.8	75.0	10.7	3.5	100.0
Of those in sector, percent placed:					
Within 50 miles	10.1	23.5	8.9	18.2	12.7
Within state	16.6	36.0	18.0	25.8	22.0

Doctoral Recipients are placed nationally, but retain regional ties



Descriptive outcomes by field

(Zolas et. al. *Science*, Dec. 2015)

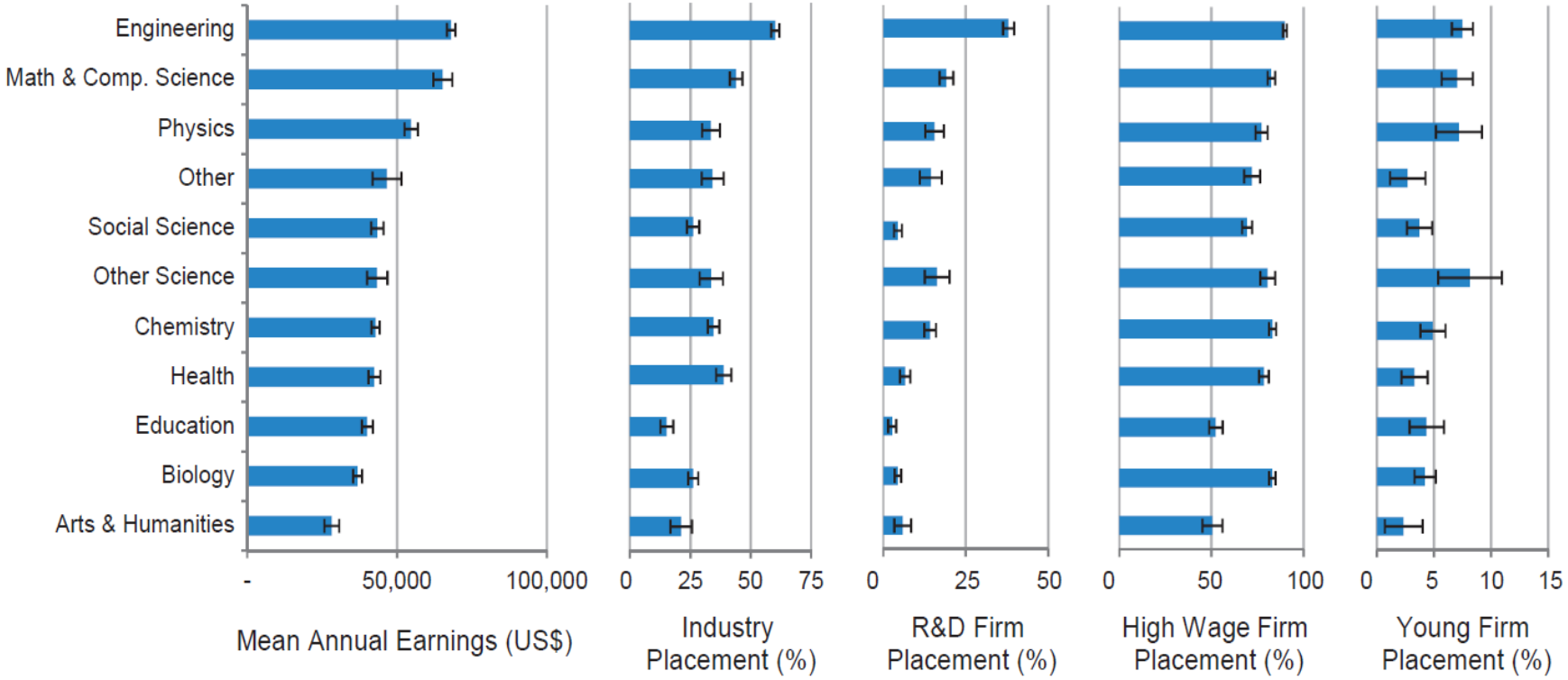
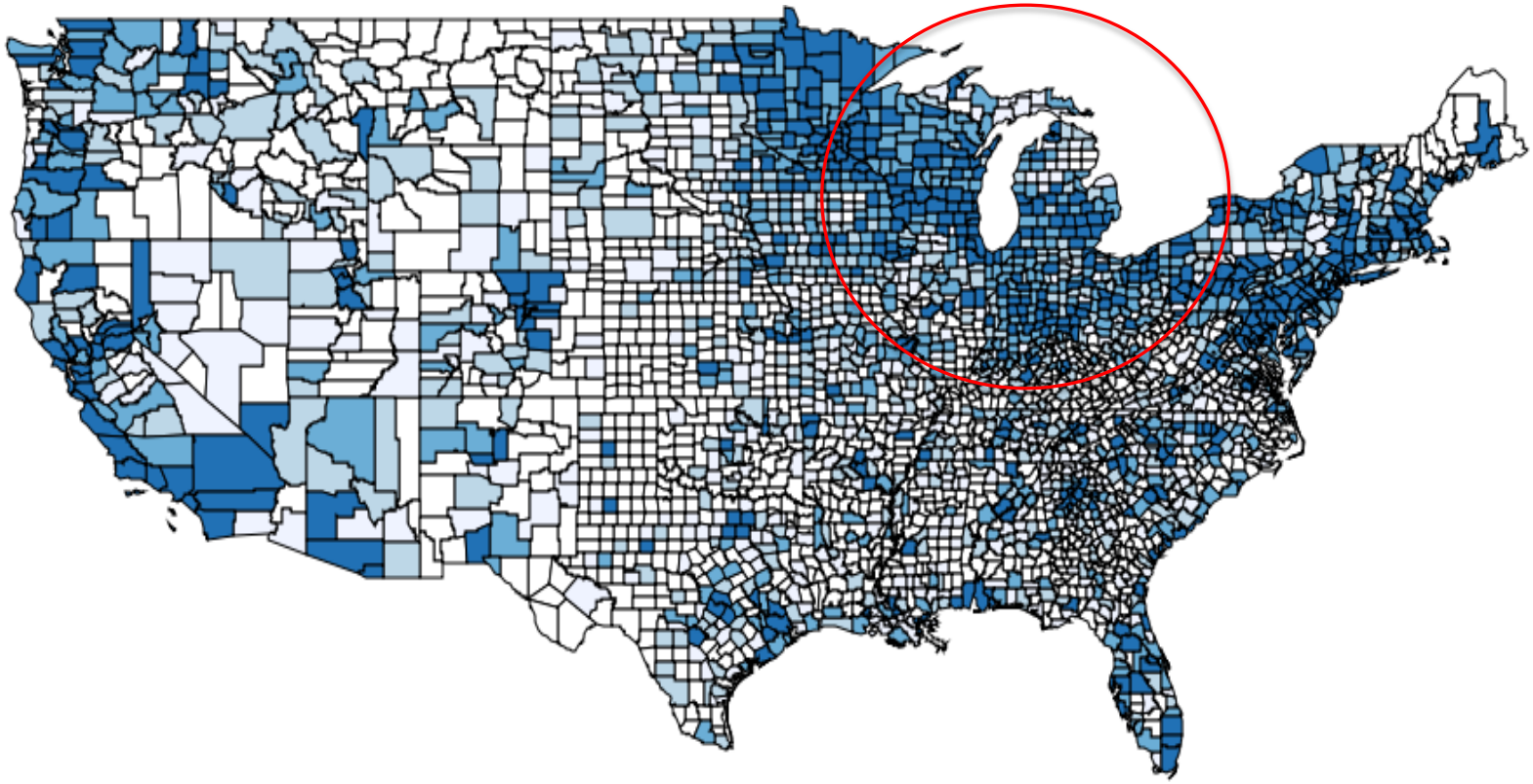


Fig. 3. The annual earnings and placement of doctoral recipients supported by grants vary by field. Young firms are defined to be those <5 years old. High–payroll per worker establishments are defined as those with a payroll per worker above the median for the establishments within their six-digit industry. Means and standard errors (error bars) for each variable.

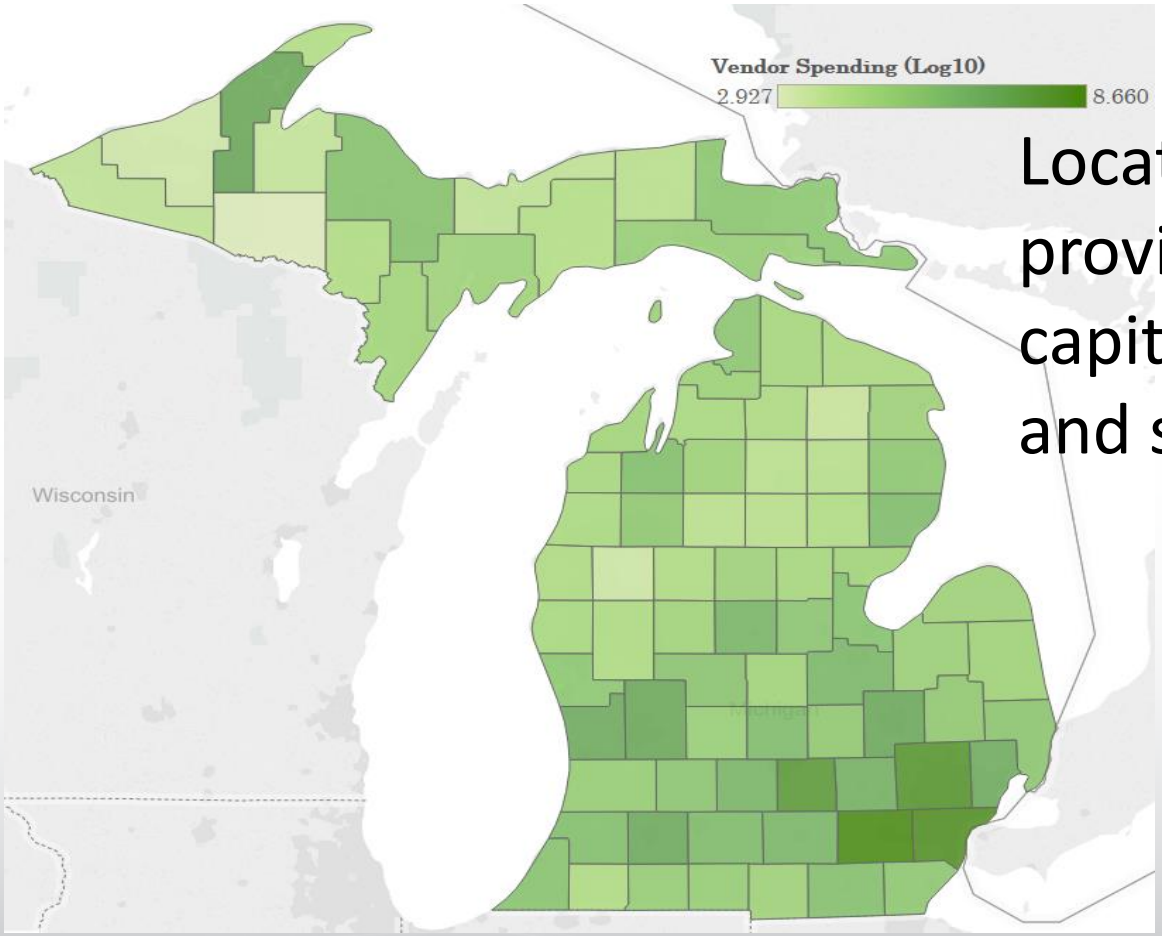
Vendor Characteristics

Geographic Distribution of Spending



Index calculated as sum of vendor payments by county for CIC universities, 2010-2012

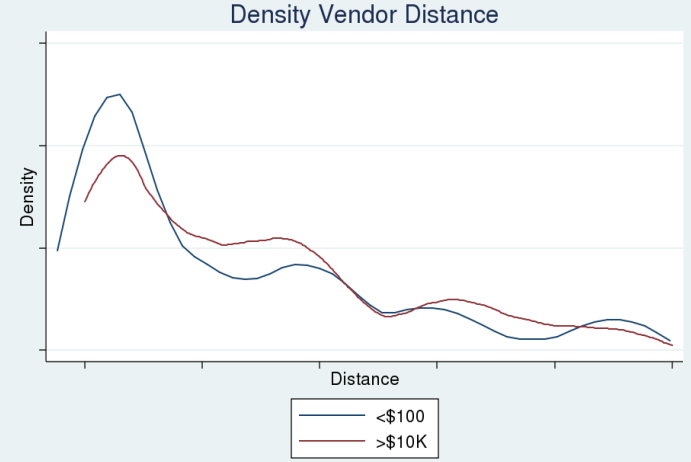
Example for universities in Michigan



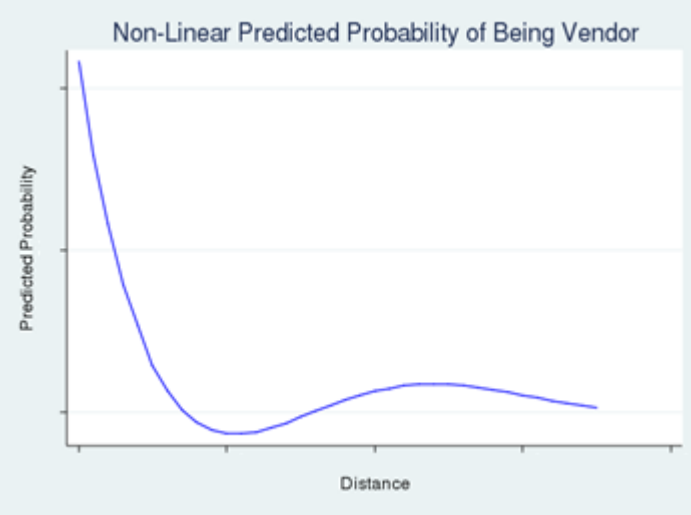
Locations of businesses providing purchased capital, materials, supplies and services

Role of Physical Distance

- The role of distance in the probability an establishment becomes a vendor is highly non-linear
  - Closer to the university, increasing distance reduces probability of being a vendor
  - Further away, effect of distance more ambiguous, maybe even positive



Mean distance and annual vendor payments between 2010 and 2012

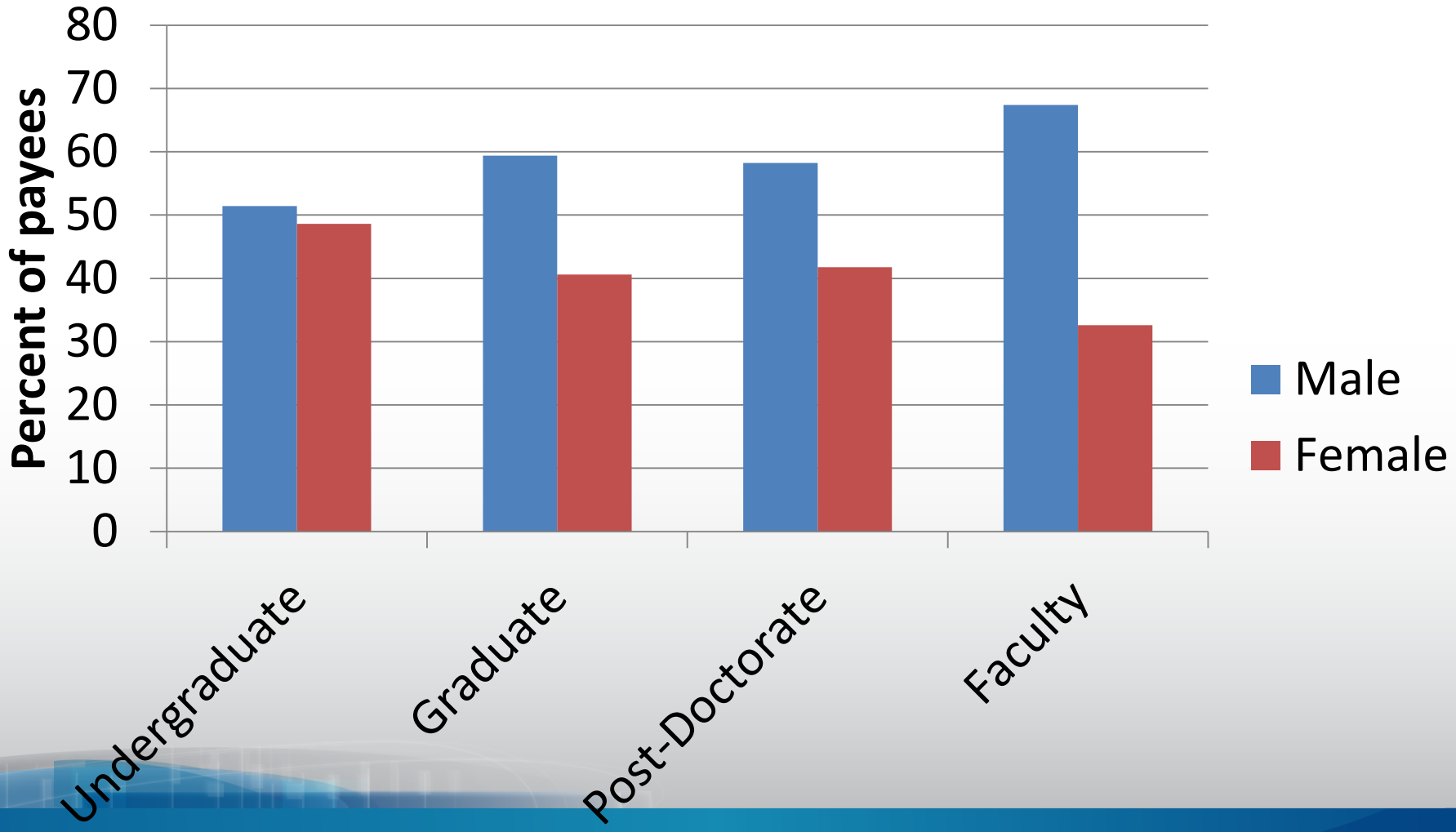


Employee Demographic Characteristics

Match to Census and ACS

2010 Census characteristics of university employees		
	Matched employee data	QWI
Percent Female	47.7	50.0
Faculty, all years	32.6	
Graduate Student, all years	40.6	
Percent Black or African American Alone	3.1	6.0
Faculty, all years	2.2	
Graduate Student, all years	3.5	
Percent Hispanic	3.9	22.1
Faculty, all years	3.0	
Graduate Student, all years	4.9	
Source: STARMETRICS employee data linked to 2010 Census, Census QWI explorer (http://qwiexplorer.ces.census.gov) 2010Q2, states with STARMETRICS institutions		

Attrition in Federal Grant Payees



Gender Wage Gap in STEM

STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census

Catherine Buffington
Benjamin Cerf
Christina Jones
Bruce A. Weinberg

AMERICAN ECONOMIC REVIEW
VOL. 106, NO. 5, MAY 2016
(pp. 333-38)

Download Full Text PDF (Complimentary)

Article Information

Abstract

Women are underrepresented in science and engineering, with the underrepresentation increasing in career stage. We analyze gender differences at critical junctures in the STEM pathway–graduate training and the early career–using UMETRICS administrative data matched to the 2010 Census and W-2s. We find strong gender separation in teams, although the effects of this are ambiguous. While no clear disadvantages exist in training environments, women earn 10% less than men once we include a wide range of controls, most notably field of study. This gap disappears once we control for women's marital status and presence of children.

The Street

Education News

Conclusions

- IRIS UMETRICS data combined with Census data provides a source of information for studying innovation and the impacts of federal research spending
- Project illustrates benefits of systematic integration of transaction and administrative data
- **These Data Offer:**
  - Employee Outcomes
    - Destination firms, employee outcomes, geographic distribution, startups
  - Vendor Attributes
  - Vendor attributes, regional impact, physical and social distance
  - Employee Demographics
- Pre-determined Products (e.g., hot reports)
- Scientific data access via the Federal Statistical System Research Data Center Network

Thank You

Ron Jarmin

Ron.S.Jarmin@census.gov