



# ProQuest Dissertations & Theses

Big Questions, Big Data

**CGS 61<sup>st</sup> Annual Meeting**

*December 2, 2021*



# INTRODUCTIONS



**Angela D'Agostino**  
VP Product  
Management



**Austin McLean**  
Senior Director of  
Academic Relations



**Gilia Smith, Ph.D.**  
Director of Academic  
Relations



**Jess Byrne**  
Senior Product Marketing  
Manager



Come visit us in the Exhibit Hall

# PQDT Brings together Scholarly Conversations

Global Community



**4.1K**

Universities represented from 60+ countries

**3.1K**

Universities subscribe to PQDT

Global Diversity



**5.2M**

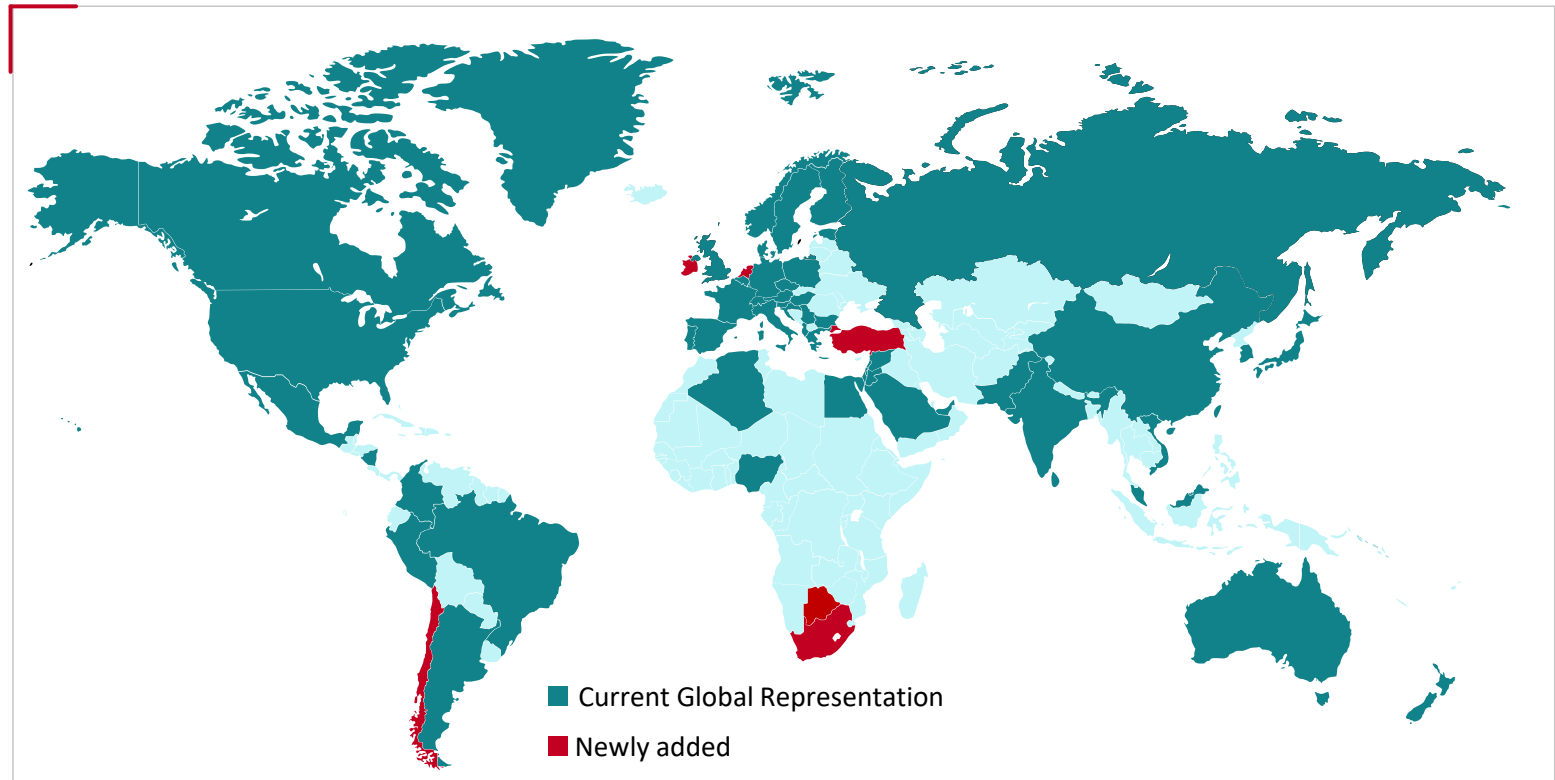
Works, including 2.8M in Full text

**200K**

Titles added annually

In 2020: **65** new universities from **23** countries joined PQDT

In 2021: **44** new universities from **19** countries joined PQDT



**Nearly 50% of Full Text added each year is international**





# ProQuest Dissertations & Theses Vision

Moving from  
Content to Insights  
to Unlock the Value  
of Graduate Student  
Research

## Three Pillars of Development



### 1 | ETD Dashboard

Gain insights on impact of your university's research



### 2 | Citation Insights

Deliver insights to researchers to improve efficiency and impact of research output

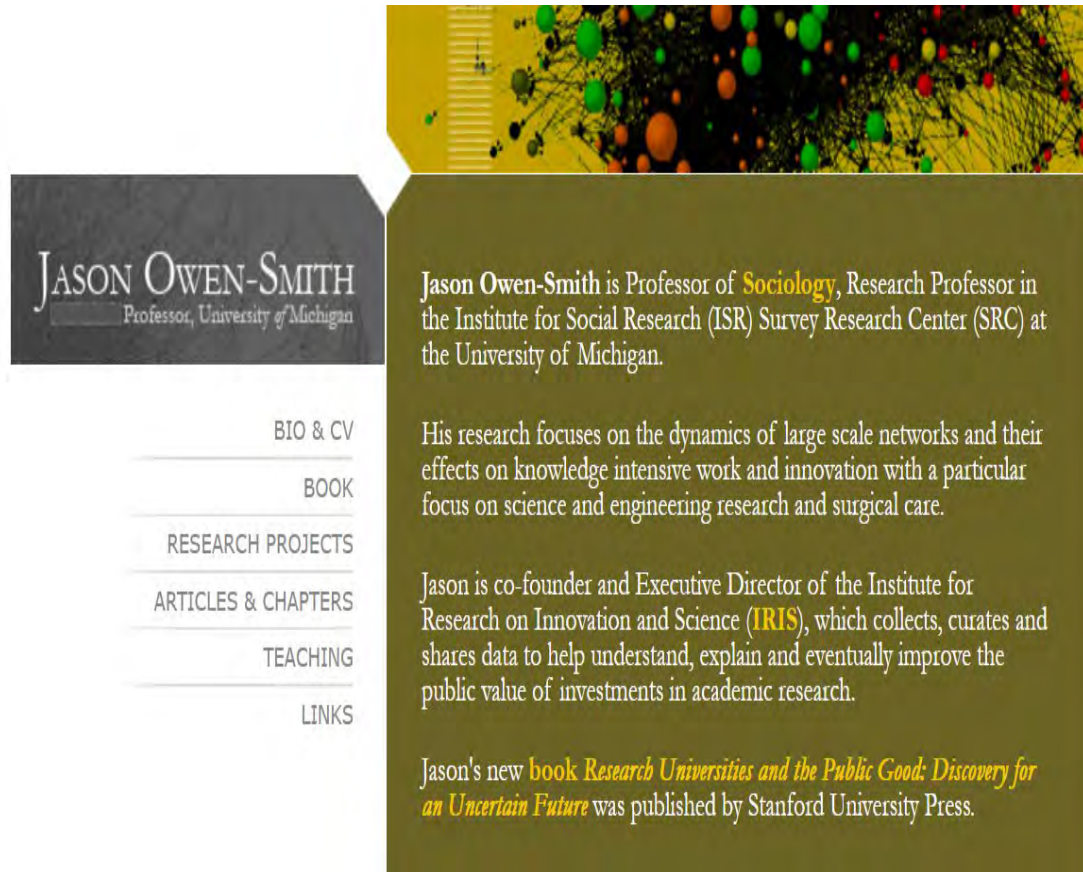


### 3 | Text and Data Mining

Extract insights from Big Data analyses



# Featured Speaker: Professor Owen-Smith



The image shows a screenshot of a website for Jason Owen-Smith. The header features a dark grey box with the name 'JASON OWEN-SMITH' in white serif font, followed by 'Professor, University of Michigan' in a smaller font. Below this is a navigation menu with links: 'BIO & CV', 'BOOK', 'RESEARCH PROJECTS', 'ARTICLES & CHAPTERS', 'TEACHING', and 'LINKS'. The main content area has a yellow background with a network diagram of green and orange nodes. The text reads: 'Jason Owen-Smith is Professor of **Sociology**, Research Professor in the Institute for Social Research (ISR) Survey Research Center (SRC) at the University of Michigan. His research focuses on the dynamics of large scale networks and their effects on knowledge intensive work and innovation with a particular focus on science and engineering research and surgical care. Jason is co-founder and Executive Director of the Institute for Research on Innovation and Science (**IRIS**), which collects, curates and shares data to help understand, explain and eventually improve the public value of investments in academic research. Jason's new book *Research Universities and the Public Good: Discovery for an Uncertain Future* was published by Stanford University Press.'

## Examples of IRIS Research using full-text documents from ProQuest Dissertations & Theses Global:

- **Measuring the Economic Value of Research: The Case of Food Safety**

Husbands Fealing, K, Lane JI, King J, & Johnson SR (eds.)

**Cambridge University Press, 2017**

- **Wrapping It Up in a Person: Examining Employment and Earnings Outcomes for Ph.D. Recipients**

Zolas N, Goldschlag N, Jarmin RS, Stephan P, Owen- Smith J, Rosen RF, McFadden Allen B, Weinberg BA, & Lane JI **Science 11 December 2015 Vol. 350(6266), pp. 1367-1371**

# Big Questions, Big Data, and ProQuest Dissertations & Theses

**Jason Owen-Smith**  
Executive Director, IRIS  
Executive Director, RADIO  
Professor, Sociology  
Research Professor, ISR  
University of Michigan  
jdos@umich.edu



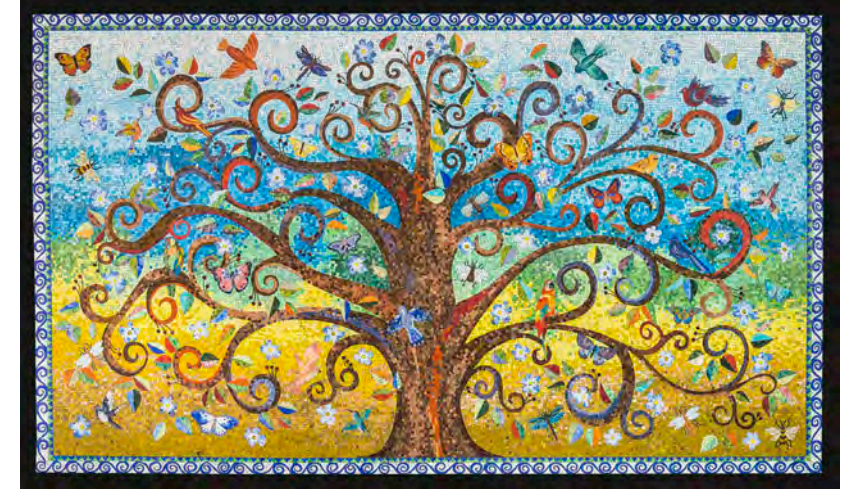
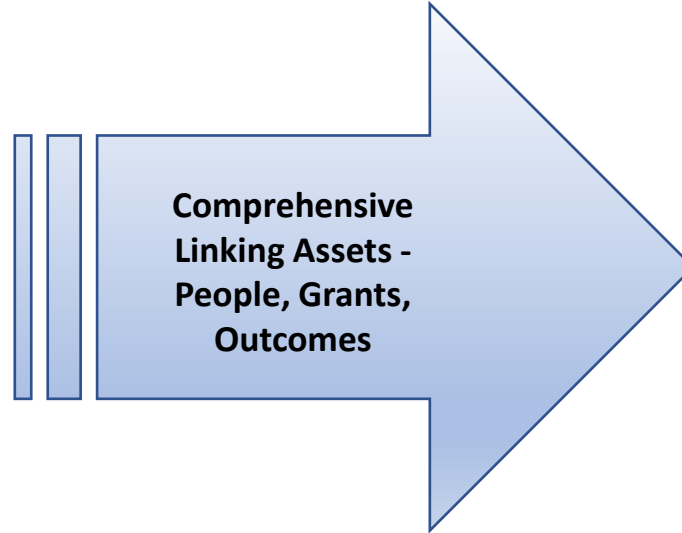
# Roadmap

- What is IRIS?
- UMETRICS data for research and reporting on graduate education
- The value of linked thesis, grant, employment & outcome data
- Examples
- How can you help?
- Discussion

# Much of the value of big data comes from constructing data mosaics



Can answer some questions  
with individual “tiles”



Exponentially more possibilities  
with linked data from many  
sources

IRIS' goal is to construct, protect, use and share a large-scale data mosaic that can answer previously unanswerable questions to help understand, explain and improve the public value of higher education and research.

# What kinds of questions can the IRIS data mosaic answer about graduate education?

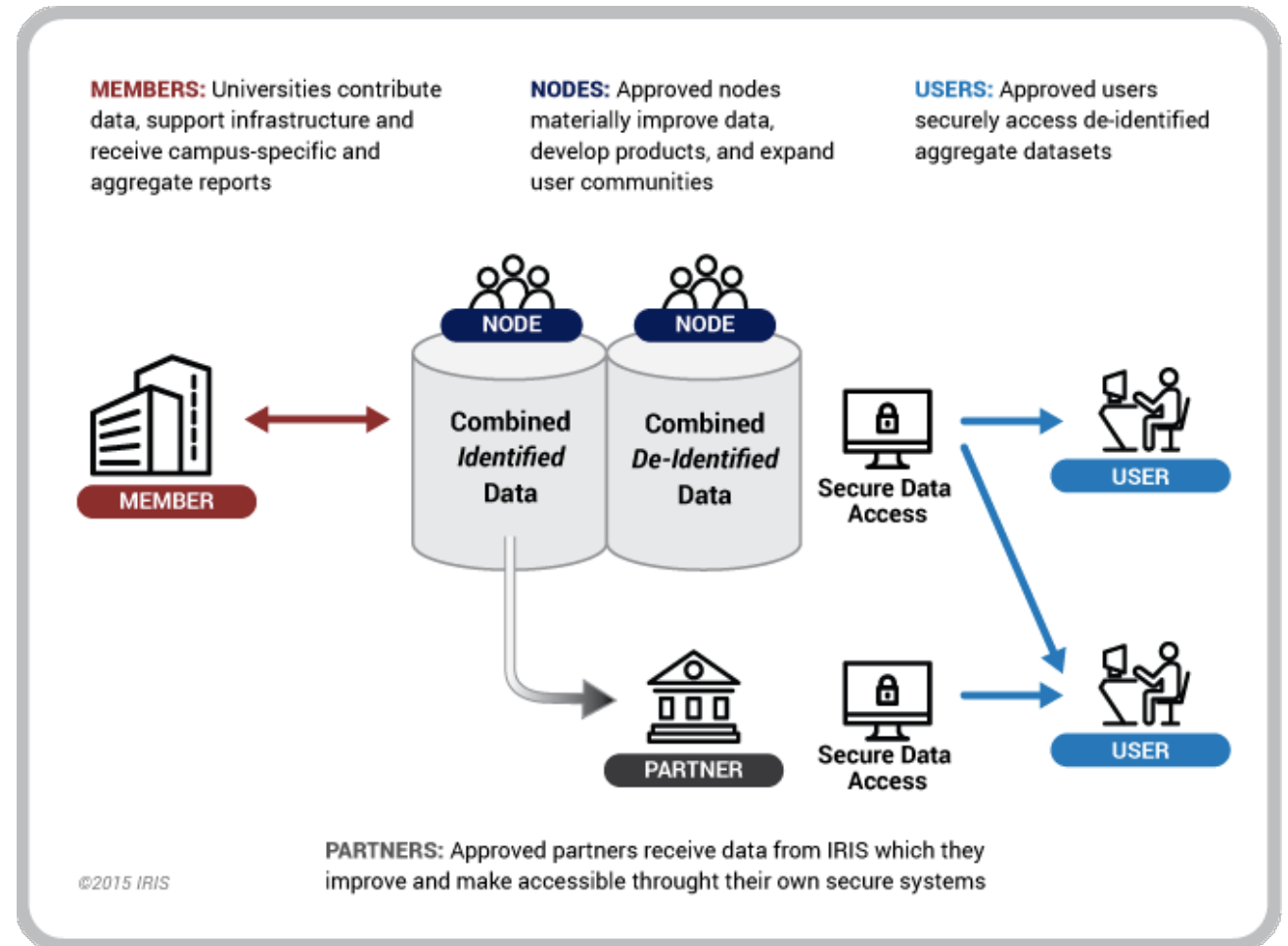
- How do we best prepare students for the full range of jobs?
- What are the best ways to support successful doctoral research?
- How do different types of research experiences/support relate to doctoral scientific & career outcomes?
- Do certain types of research training develop hard and soft skills more effectively?
- How does grad school debt, student demographics, and different types of research training create and sustain inequities?
- ...

By creating, protecting and sharing comprehensive data mosaics and helping researchers in all fields develop the capability to use them effectively and responsibly we ensure that these questions are addressed rigorously and replicably by a diverse, interdisciplinary research community.

# How Does IRIS Work?

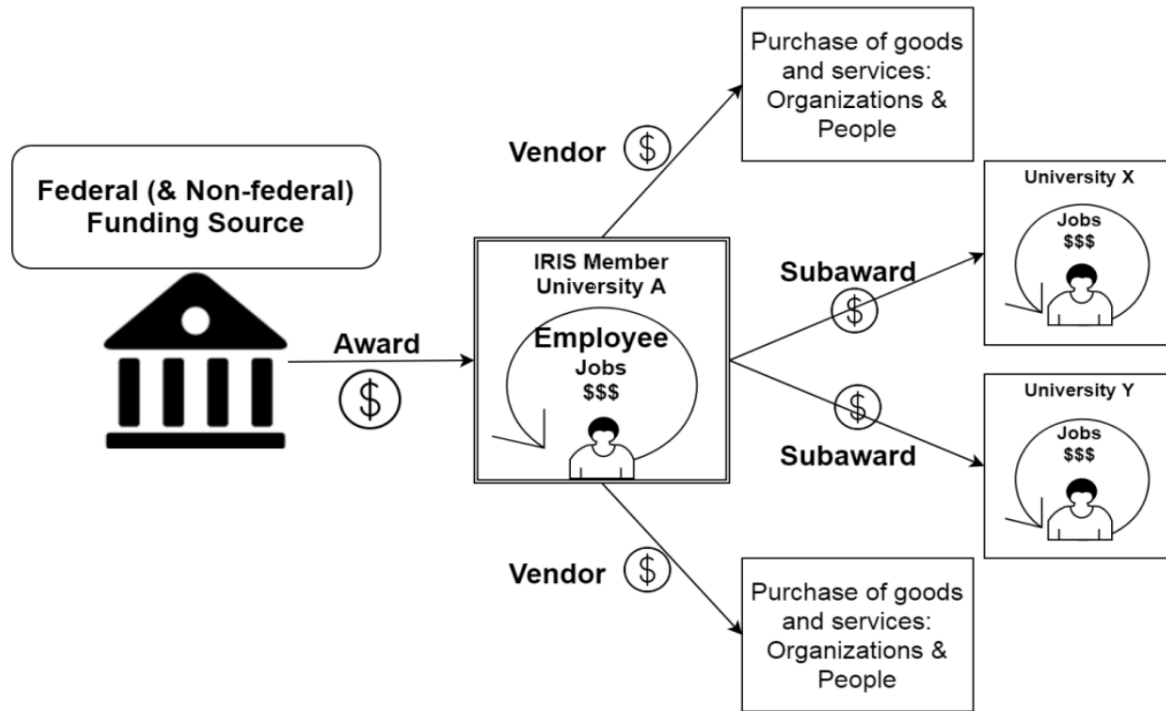
## Currently

- Data on 450,000 (federal & non-federal) sponsored projects that pay 721,000 people
- Data on ~\$100 billion of research spending
- Broadly representative of NSF & NIH Award Portfolios
- Data on >41% Academic R&D spending
- 4 research data releases
- >340 users from about 150 institutions





# UMETRICS 2020 Dataset



**Integrate & Link with > 50 data streams**

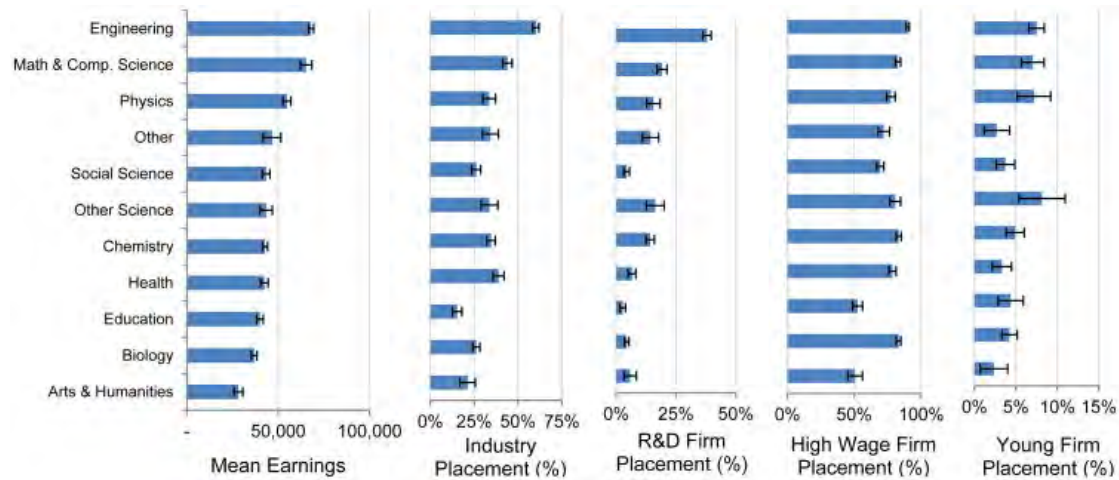
Of relevance to questions about Grad Students

- *Research Characteristics & Outcomes*– ProQuest Dissertations, Medline, Web of Science, Federal Grants
- *Employment Outcomes* – SteppingBlocks, Census Bureau, SED
- *Debt & Graduate School Experience* – SED/SDR, competition measures, funding types and intensities, time to degree, inequities

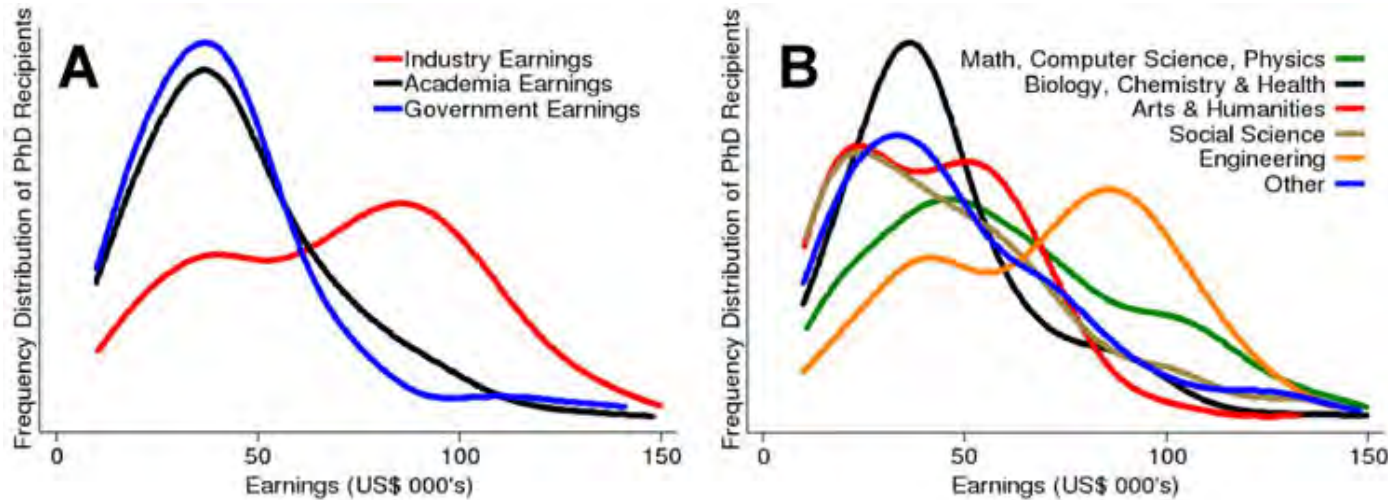
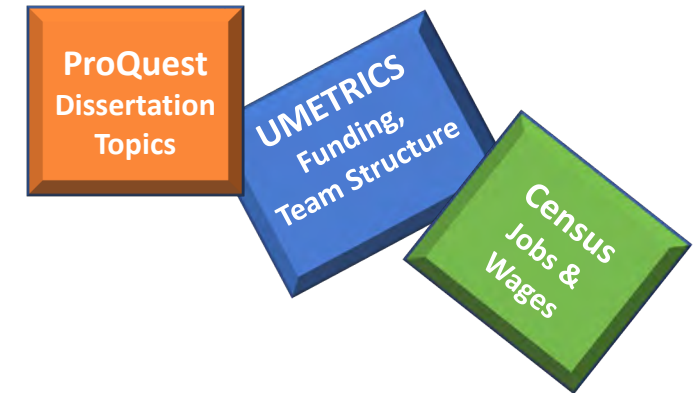
**Data on: > 166,000 Doctoral Students, > 140,000 Master's Students**

For understanding graduate outcomes, comprehensive dissertation data that includes summary features (e.g. well defined topic ontologies) metadata (e.g. advisors and committee members, departments) and text (e.g. abstracts) is a key component

# Example 1: Academic & Non-Academic Outcomes



Essential “Tiles”



# Example 2: The tension between research productivity & graduate training

## Rescuing US biomedical research from its systemic flaws

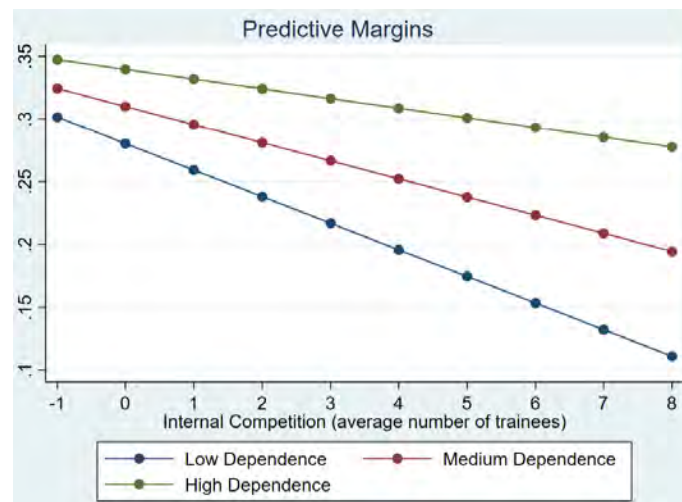
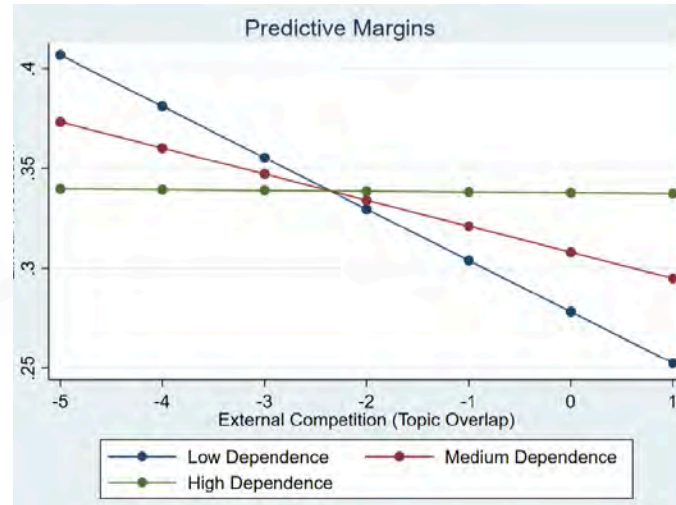
Bruce Alberts<sup>a</sup>, Marc W. Kirschner<sup>b</sup>, Shirley Tilghman<sup>c,1</sup>, and Harold Varmus<sup>d</sup>

<sup>a</sup>Department of Biophysics and Biochemistry, University of California, San Francisco, CA 94158; <sup>b</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115; <sup>c</sup>Department of Molecular Biology, Princeton University, Princeton, NJ 08540; and <sup>d</sup>National Cancer Institute, Bethesda, MD 20892

Edited by Inder M. Verma, The Salk Institute for Biological Studies, La Jolla, CA, and approved March 18, 2014 (received for review March 7, 2014)

The long-held but erroneous assumption of never-ending rapid growth in biomedical science has created an unsustainable hypercompetitive system that is discouraging even the most outstanding prospective students from entering our profession—and making it difficult for seasoned investigators to produce their best work. This is a recipe for long-term decline, and the problems cannot be solved with simplistic approaches. Instead, it is time to confront the dangers at hand and rethink some fundamental features of the US biomedical research ecosystem.

graduate education | postdoctoral education | federal funding | peer review



Essential “Tiles”

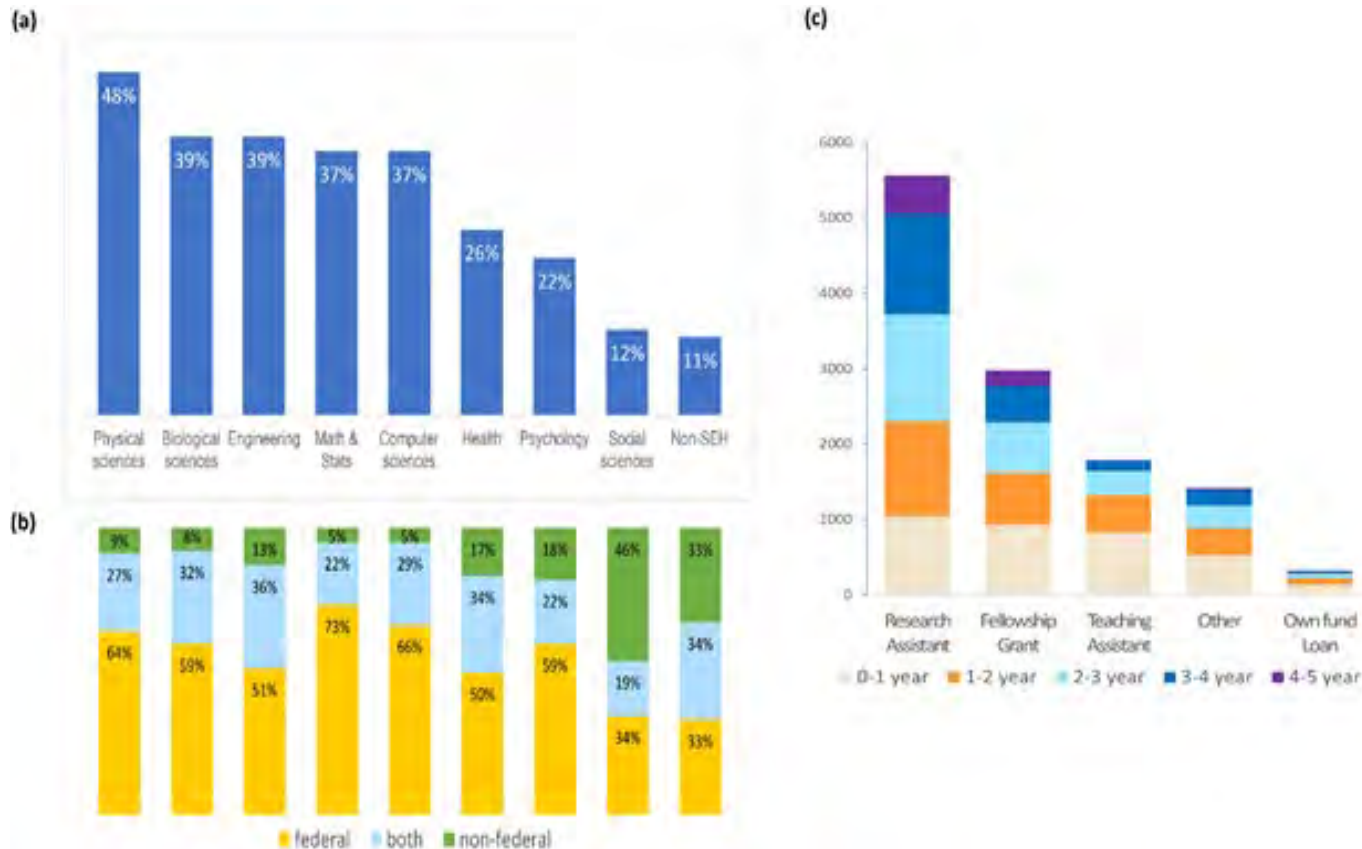


Source: Oh & Owen-Smith, 2021

# Example 3: Debt, equity and the paradox of diversity

**The Paradox:** Diverse research teams produce better science, but the people who make teams diverse have worse individual outcomes

**Possible Mechanisms:** greater dependence, uneven costs of coordination, inequitable credit, debt constraints, access to research funding



Essential “Tiles”





# Next Steps: Expanding Data, Access, Research Community Capacity

- Expanding NCSES partnership
- Growing IRIS
  - Scale: More universities
  - Scope: Student enrollment data and directory data (in pilot)
- Integrating behavioral surveys of hard/soft skill development (with David Feldon)
- Integrating dissertations and theses from new universities into the ProQuest database
- Education & Social Science Community Capacity Building
- Developing open access publication outcome data
- Expanding privacy protection tools

# What can you do?

- Help us bring your university into IRIS
- Maintain comprehensive ProQuest dissertation and theses data
- If your campus is part of IRIS, participate in skill surveys
- Spread the word to researchers on your campus

# Talk with me about IRIS Membership

## Benefits

- 9 campus specific data products: government relations/communications, student and post-doc employment outcomes, economic impact
- Your student and faculty researchers get free, secure access to UMETRICS research data and research support through a virtual data enclave
- Your campus gets free access to your improved data and UMETRICS research data for IR and evaluation purposes
- Your campus can participate in governance and have a seat at the table on product and data development
- Your campus is represented in IRIS' aggregate data reports for the public, policy makers, agencies etc.

## Costs

- Contribute \$25,000/year to support IRIS infrastructure
- Sign IRIS Membership MOU
- Two points of contact
- Produce and submit data annually (~1-2 weeks analyst time year 2, negligible in subsequent)



## Contact me via

- Email ([jdos@umich.edu](mailto:jdos@umich.edu))
- Text (734-276-8923)
- CGS Meeting App

to find time to talk in New Orleans  
or after

Thoughts about IRIS or ProQuest?