

The University Libraries as Data Science Research Partner

Dr. Bruce Herbert, Director
Office of Scholarly Communications

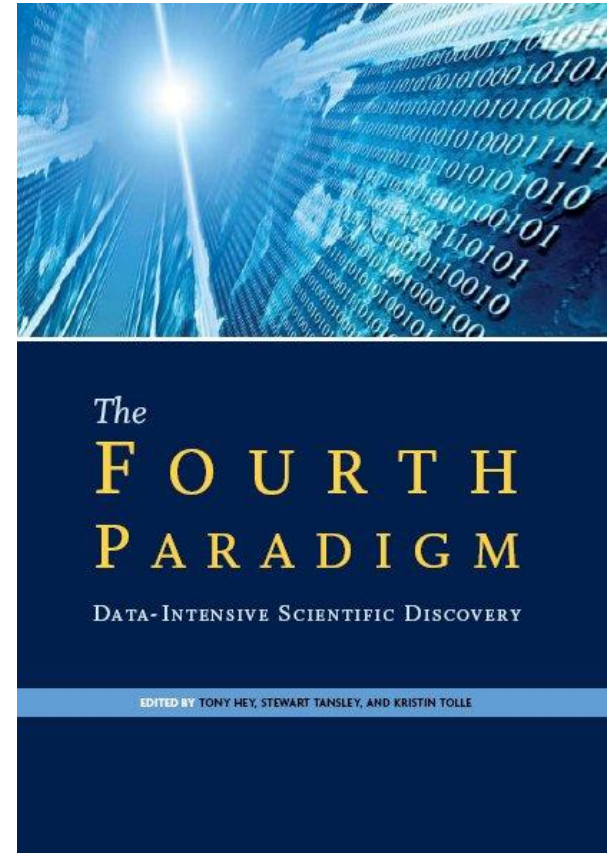
The Fourth Paradigm: Data-Intensive Scientific Discovery

“In Jim Gray’s last talk to the Computer Science and Telecommunications Board on January 11, 2007, he described his vision of the fourth paradigm of scientific research.

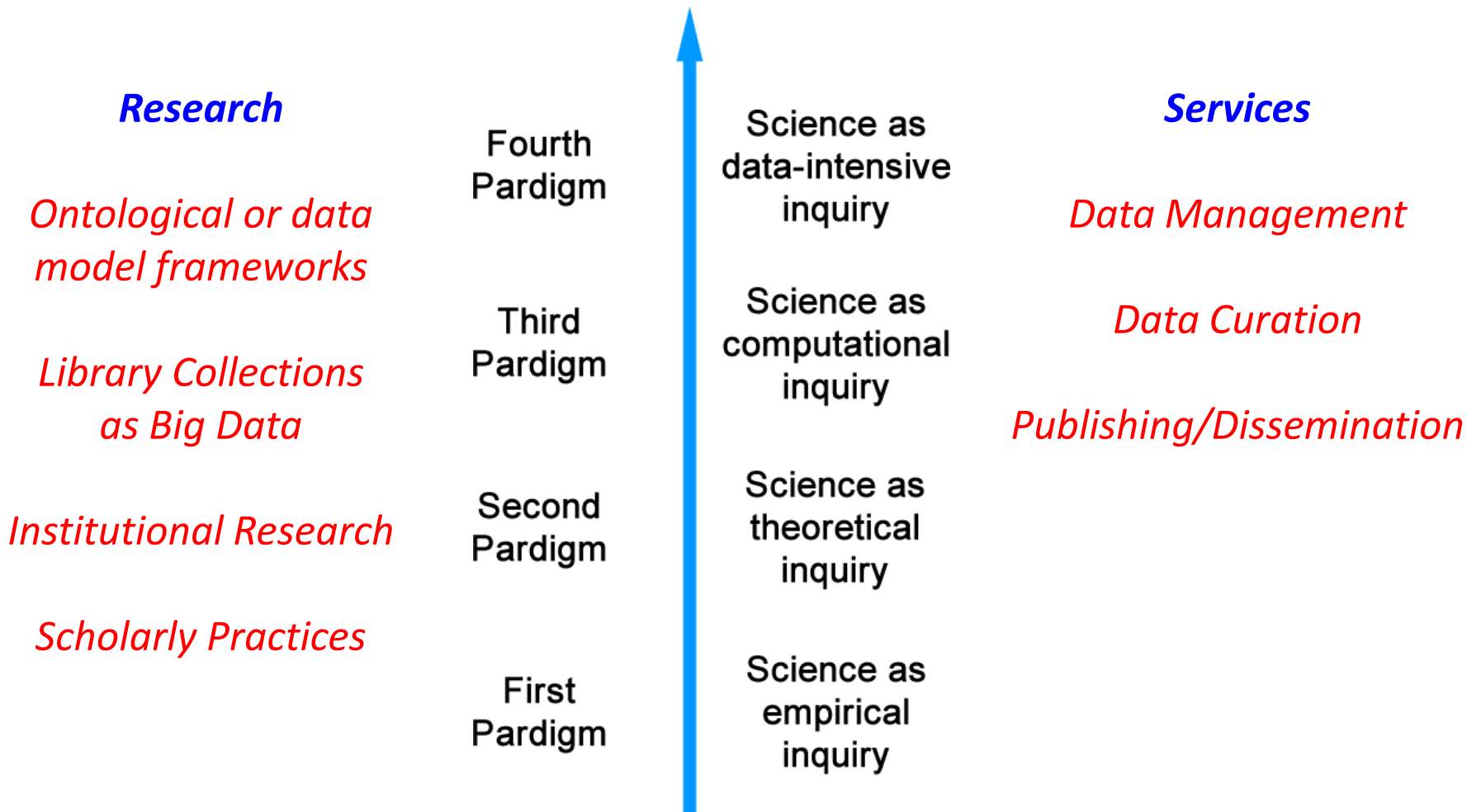
He outlined a two-part plea for the funding of tools for data capture, curation, and analysis, and for a communication and publication infrastructure.

He argued for the establishment of modern stores for data and documents that are on par with traditional libraries.”

<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>



A Transformed Scientific Method



University Libraries as:

Service Provider

Source of Big Data

Research Partner

Federal Mandates For Public Access to Research

the WHITE HOUSE PRESIDENT BARACK OBAMA

Get Email Updates | Contact Us

BLOG PHOTOS & VIDEO BRIEFING ROOM ISSUES the ADMINISTRATION the WHITE HOUSE our GOVERNMENT

Home • The Administration • Office of Science and Technology Policy

Search WhiteHouse.gov

Office of Science and Technology Policy

About OSTP | Pressroom | OSTP Blog | Divisions | Initiatives | R&D Budgets | Resource Library | NSTC | PCAST | Contact Us

Expanding Public Access to the Results of Federally Funded Research

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST

E-Mail Tweet Share +

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand of whom recently signed a *We the People* petition asking for expanded public access to the results of taxpayer-funded research.

To see the new policy memorandum, please visit: http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

To see Dr. Holdren's response to the *We the People* petition, please visit: <https://petitions.whitehouse.gov/response/increasing-public-access-results-scientific-research>

Michael Stebbins is Assistant Director for Biotechnology at OSTP

The Library Supports:

Publication repositories

Training

Funding for OA
publication fees

Data repositories

<http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>

Research Innovation: Data Management

Data is citable

Meets funder mandates

Supports move to open science

Link graduate student data to ETDs in TAMU repository

Library training programs

The screenshot shows the Texas Data Repository website. At the top, there is a navigation bar with the logo and search options. Below that, a banner reads "Texas Data Repository Dataverse" and "A statewide collaboration of Texas higher education institutions". A metrics box shows "152 Downloads". The main heading is "Share, publish, and archive your data. Find and cite data across all research fields." Below this is a welcome message and a "LEARN MORE" section with links to the user guide and local university librarians. A carousel of logos for member institutions is shown: Texas State University, Texas A&M University, Angelo State University, and Baylor University. A search bar is present with "Find" and "Advanced Search" buttons. The search results section shows "1 to 10 of 89 Results" and lists two entries: "JGR BernardBehr2017 FTIR data" and "JGR BernardBehr2017 EBSD data", both dated August 13, 2017, and authored by Rachel Bernard. Each entry includes a DOI and a brief description of the data files.

<https://data.tdl.org/>

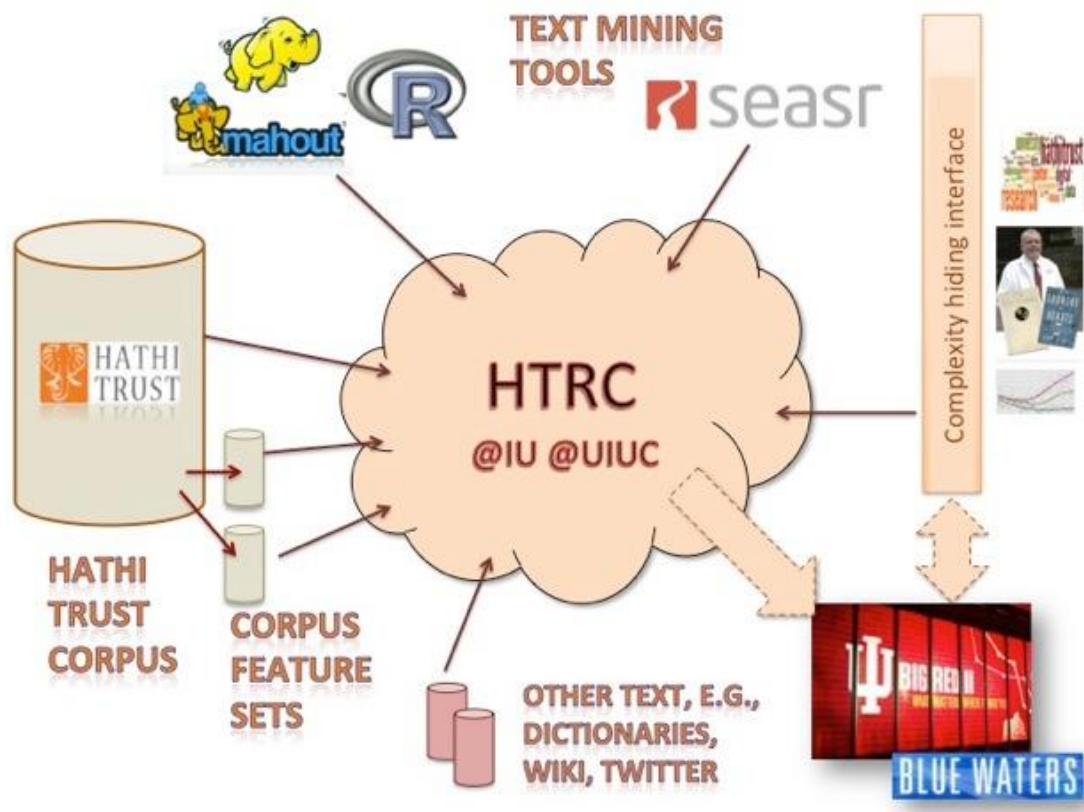
University Libraries as:

Service Provider

Source of Big Data

Research Partner

Library Collections as Data: Text Mining Library Collections

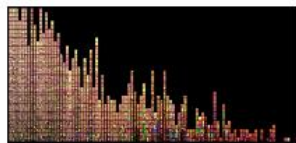


Library Collections as Data: Text Mining Library Collections



Few magazines can boast being continuously published for over a century, familiar and interesting to almost everyone, full of iconic pictures — and also completely digitized and marked up as both text and images. What can you do with over 2,700 covers, 400,000 pages, 6 TB of data? Students, librarians and faculty are excited about the possibilities of working with *Vogue* to explore questions in fields from gender studies to computer science. We highlight some early experiments below:

Slice Histograms



Direct visualization of color patterns.

Explore »

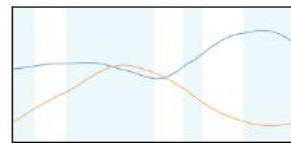
Cover Averages



Visual continuity and change across the decades.

Explore »

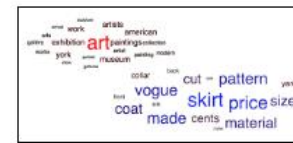
n-gram Search



Search and compare word usage in 400,000+ pages within ads, articles, or all texts.

Explore »

Topic Modeling



Self-organizing themes determined by word co-occurrence.

Explore »

Advertisements



Sort ads by frequency, date, and industry.

Explore »

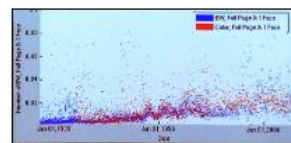
Statistics



View figures on circulation, ratio of articles to advertisements, price per issue, and number of pages per year.

Explore »

Student Work



See projects conducted by Yale students using this data.

Explore »

Colormetric Space



Hue, Saturation and Lightness.

Explore »

University Libraries as:

Service Provider
Source of Big Data
Research Partner

[Hours](#)
[Libraries](#)
[MyLibrary](#)
[Help](#)

SCHOLARS@TAMU

[Home](#)
[People](#)
[Organizations](#)
[Research](#)
[About](#)
[Update Profile](#)

Statistics

- 2.6^k People
- 3.5^k Organizations
- 169^k Research

Search Scholars@TAMU

[Search](#)

Thank you for visiting Scholars@TAMU (Beta Release). We have worked to create the first version of your profile and continue to explore additional data sources to enhance your profile. We request your participation to update your profile and improve its accuracy. If you have any concerns or questions about profile page, please contact us at scholars@library.tamu.edu.

[Explore](#)
[Update](#)
[About](#)

[Find People](#)
[Profile Information](#)
[New to Scholars@TAMU?](#)

[Browse TAMU Organizations](#)
[Edit My Profile](#)
[Know Your Impact](#)

[Discover TAMU Research](#)
[Issues with Profile](#)
[FAQ](#)

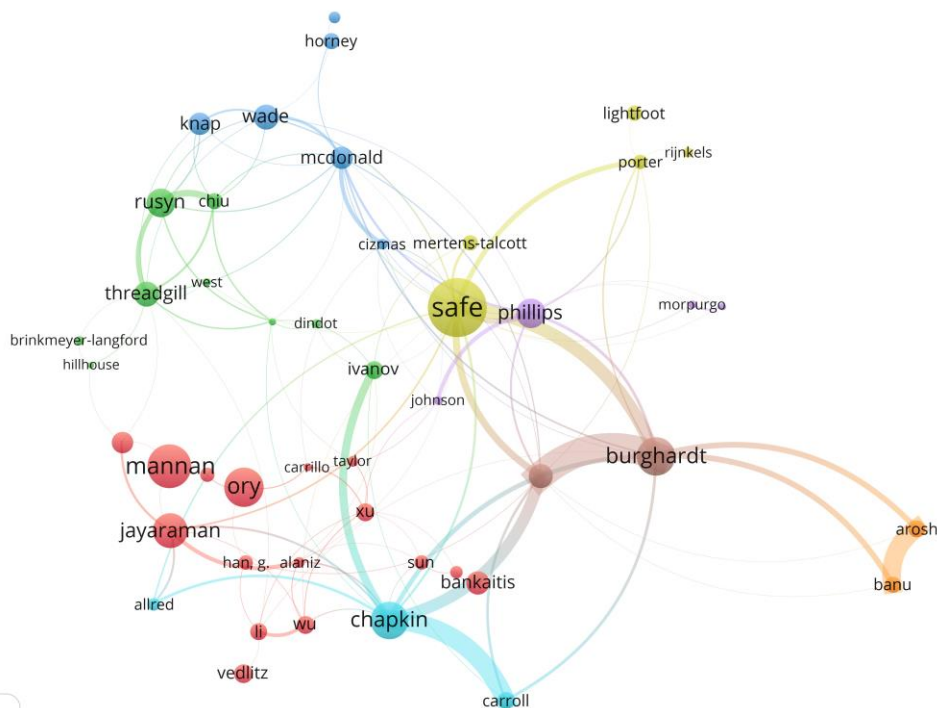
[Link my ORCID](#)
[Contact Us](#)

Scholars@TAMU

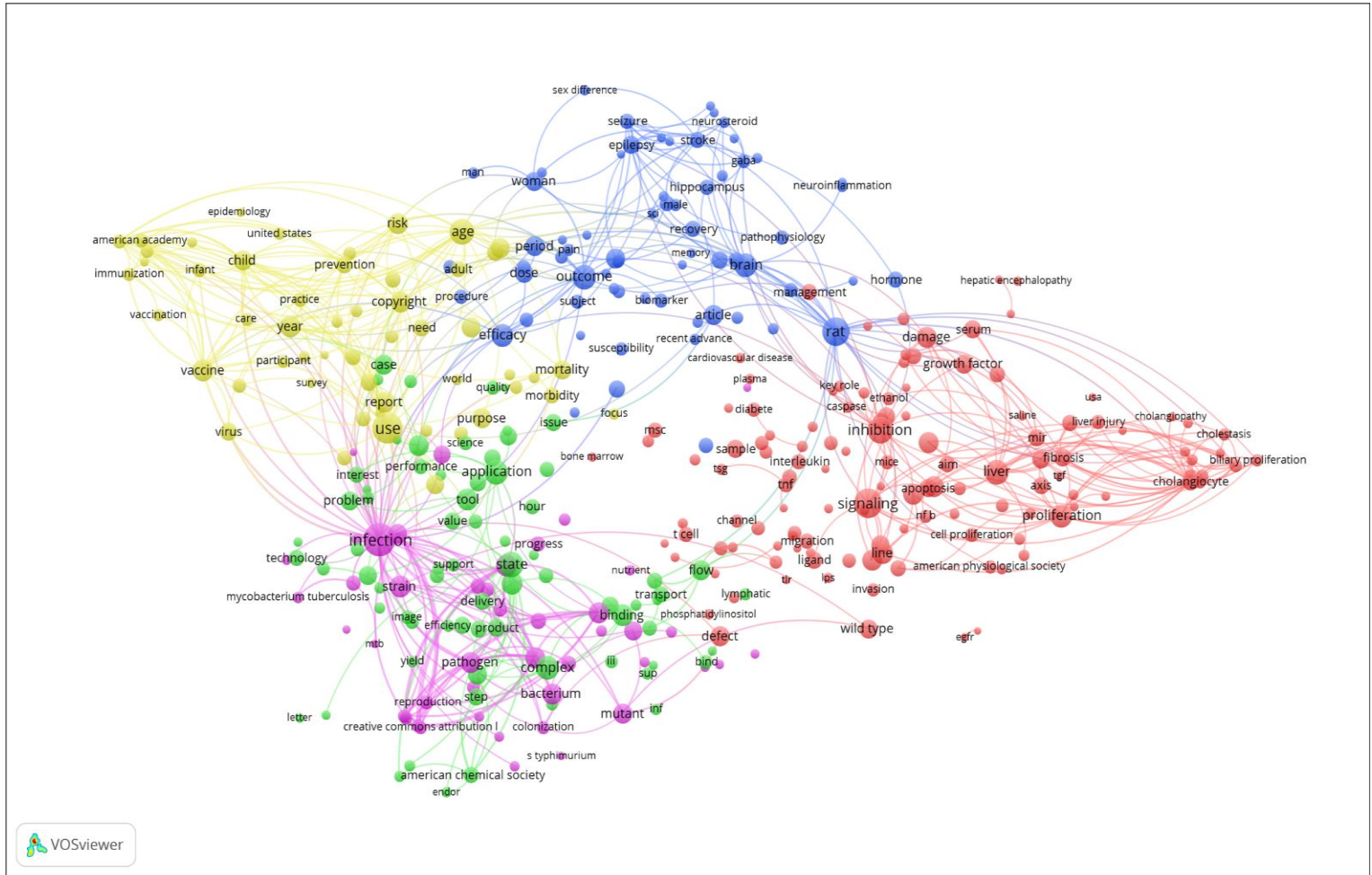
- Integrated Impact Metrics
- Publications
- Grants and Awards
- Teaching
- Thesis & Dissertations

<http://scholars.tamu.edu/>

Institutional Research



Research Strengths
Scholarly Impact & Reputation
Collaboration Networks
Societal Impact



Network visualization of research areas in the College of Medicine

Impact of Team Collaboration

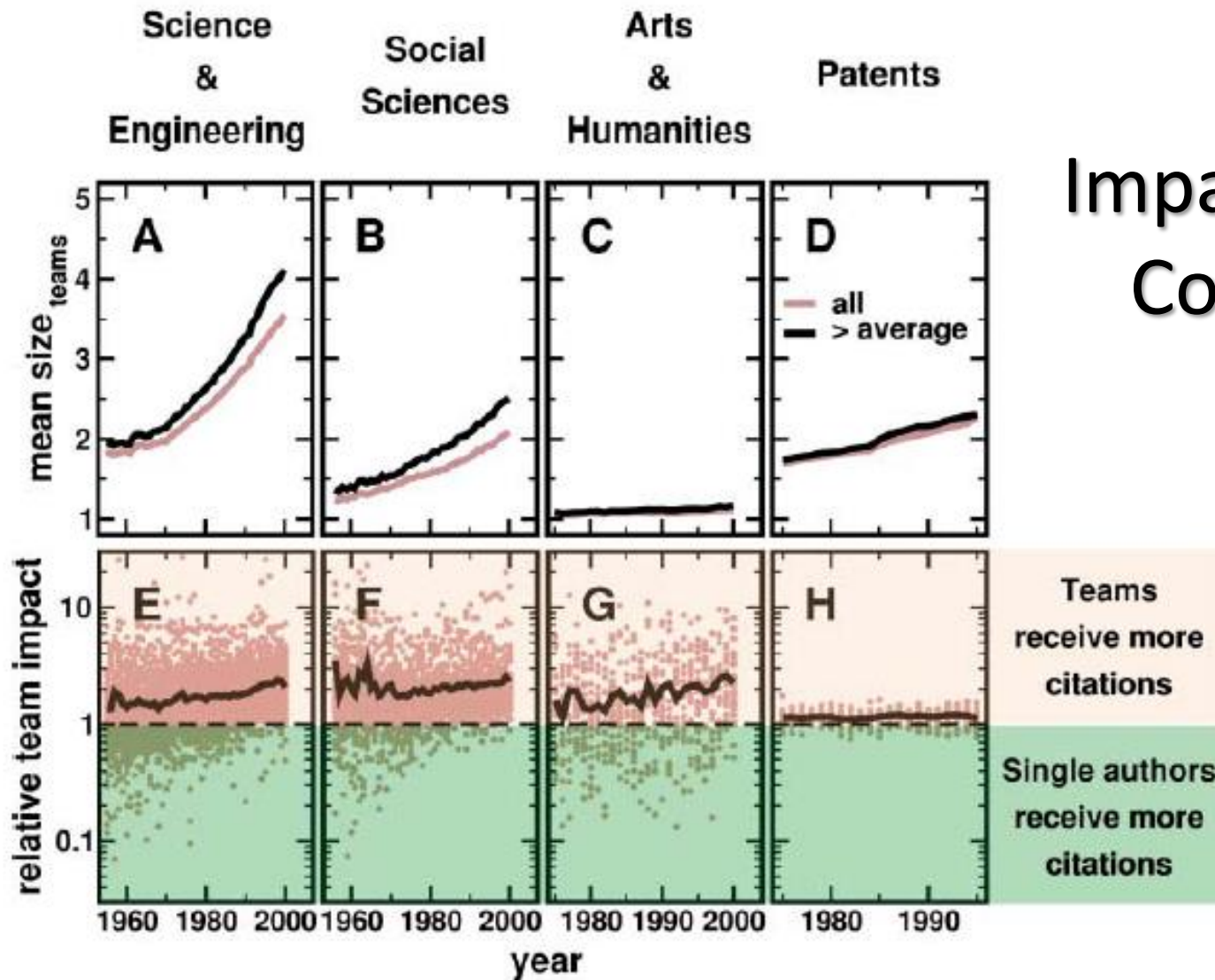


Fig. 2. The relative impact of teams. (A to D) Mean team size comparing all papers and patents with those that received more citations than average in the relevant subfield. (E to H) The RTI, which is the mean number of citations received by team-authored work divided by the mean number of citations received by solo-authored work. A ratio of 1 indicates that team- and solo-authored work have equivalent impact on average. Each point represents the RTI for a given subfield and year, whereas the black lines present the arithmetic average in a given year.

Wuchty et al. (2007).
Science 316(5827):
1036-1039.

Collaboration Across Institutions

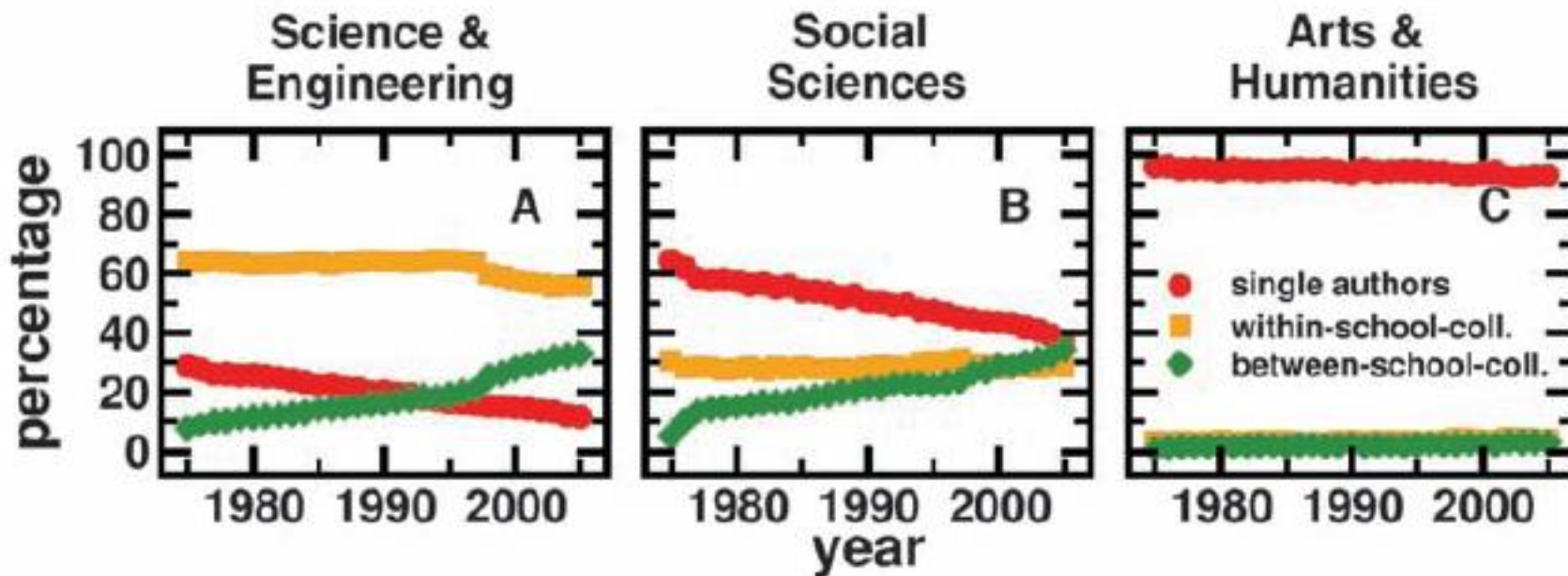


Fig. 1. The rise in multi-university collaboration. By comparing the incidence of papers produced by different authorship structures, we see that the share of multi-university collaborations strongly increases from 1975 to 2005. This rise is especially strong in SE (A) and SS (B), whereas it appears weakly in AH (C), in which collaboration of any kind is rare. The share of single-university collaborations remains roughly constant with time, whereas the share of solo-authored papers strongly declines in SE and SS.

Jones et al. (2008). *Science* 322: 1259-1262.

University Libraries as:

Service Provider
Source of Big Data
Research Partner

Dr. Bruce Herbert
beherbert@tamu.edu