

Discussant Comments

Big Data

Juan Pablo Hourcade
University of Iowa

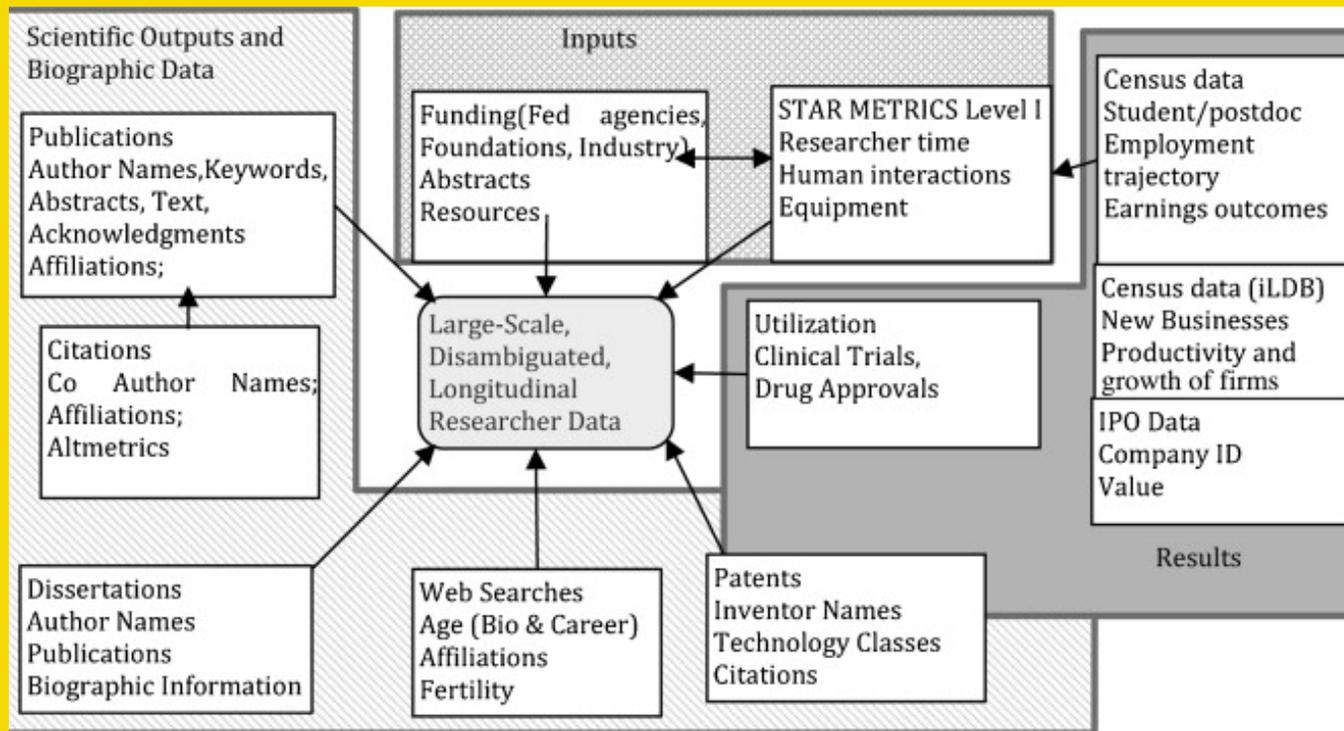
Overview

- **Need** for greater frequency and granularity of data
- **Opportunity** to leverage external data sources

Activities

- Research on big data sources, methods, and procedures
 - Specific goal of leveraging third-party data sources
 - IRIS, NPD, Retail Trade survey data
 - Web scraping
 - Public Sector statistics
 - APIs (web services?)
 - Building permit data

IRIS data



IRIS data

- Promising beginning
- Greatest challenge would be to follow innovations when they don't follow people directly
 - estimate economic impact of innovation

NPD

- *“The world’s most successful brands turn to The NPD Group to know what is selling, who is buying, and why.”*
- Challenge of mismatches in data (lack of granularity)
- Would be interesting to investigate differences in trends

Retail third-party data

- Large amount of detailed data
- Common sense projects being pursued
- Can this data source be leveraged to expand/support current data products?

Web scraper for public sector statistics

- May be challenging, depending on how data is presented
- Take advantage of site-specific standards
 - Hybrid models with human input?
- Try deep learning approaches
 - Model high-level abstractions in data through the use of multiple processing layers

APIs for building permit data

- Common sense approach
- Lack of standardization
 - Different legal requirements per jurisdiction
 - What could be standardized?

Questions

- Fay-Herriot models for small area estimation
 - Better answered by entire committee
- Additional data sources
 - Better answered by entire committee

Reducing respondent burden

- Lower barriers (and steps) to contributing data
- Consider making respondent effort proportionate to reward
- Provide immediate rewards to respondents
 - and continue communication when more data is available

Reducing respondent burden

- Communicate with states/local governments/vendors to establish common standards for web services/APIs
- Scraping can provide short term solutions
- Strategic purchasing/partnerships on datasets

Data blending

- Two-way street
 - good start <https://www.census.gov/developers/>
- Transparency
- Quality control
- Use different branding to distinguish from products with no blending