



Using Bibliographic and Census Administrative Databases to Identify FSRDC Research Outputs

Laura Fang

Supervised by Heather Madray & David Beede
U.S. Census Bureau, Center for Enterprise Dissemination (CED)

An Information Gap



- Lack of material for demonstrating usefulness of data-sharing program
- No efficient way of finding publications

Problem Objectives:

- Use Python to automate a process to collect publication data from bibliographic databases (Base, Core, etc.)
- Use NLP and publication metadata to identify possible hits (matches between Census projects and published works)

Researchers

Publication Year

Title keywords

Dataset names

Abstract phrases

Benefits

Almost fully automated process to reduce resource needs

Will allow information to be updated more regularly (projected quarterly vs. annually basis)

Narrows down possible matches from thousands to tens

Publication data shows usefulness of Census program

Provide examples of how restricted-access Census Bureau data can be used for research

Next Steps

- Commenting and creating documentation for easy transfer of work
- Collecting human “data” for ML training
- Comparing results across databases for further refinement

Deliverable:

Webpage matching
Census-approved projects
to the bibliographical
information of their
resulting publications