U.S. Census Bureau Innovation Measurement Initiative (IMI)
and
Institute for Research on Innovation & Science (IRIS):

2018 Research Data Release

June 7, 2018

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

IRIS | INSTITUTE FOR RESEARCH ON INNOVATION & SCIENCE

FUTURE ON
Activating Change.

Today's presenters:

Shawn Klimek
Assistant Center Chief for Research
Center for Economic Studies
U.S. Census Bureau

Topic:
Census and Innovation Measurement Indicators (IMI)

Jason Owen-Smith
Executive Director
Institute for Research on Innovation
and Science (IRIS)

Topic:
IRIS Overview

Barbara Downs
Director, FSRDC
Center for Economic Studies
U.S. Census Bureau

Topic:
Using Census Data at the Federal Statistical Research
Data Centers

FUTURE ON
Activating Change.

Census and Innovation Measurement Indicators (IMI)

Shawn Klimek
Assistant Center Chief for Research
Center for Economic Studies
June 7, 2018

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

3

FUTURE ON
Activating Change.

Census Bureau

- The Census Bureau's *mission* is to serve as the leading source of quality data about the nation's people and economy. We honor privacy, protect confidentiality, share our expertise globally, and conduct our work openly.

- We are guided on this mission by scientific objectivity, our strong and capable workforce, our devotion to research-based innovation, and our abiding commitment to our customers.

FUTURE ON
Activating Change.

Center for Economic Studies

- In the Research and Methodology Directorate (ADRM)

- The Center for Economic Studies (CES) <u>partners with stakeholders within and outside the Census Bureau</u> to improve measures of the economy and people of the United States through research and the development of innovative data products.

- Research Area
  - Business Economic Research Group (LBD, ILBD, LFTTD; BDS)
  - LEHD Economic Research Group (LEHD Infrastructure; QWI, LODES, J2J)
  - Research Data Centers (12 in 2012, 30 today)

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

5

FUTURE ON
Activating Change.

IMI History…

- Initially IMI was supported directly by ADRM (Jarmin)…
  - Big Data Class/Training was a big component…
  - Current and former CES staff (Buffington, Cerf, Goldschlag, Zolas) helped build out the infrastructure
- Later, Center for Big Data Research and Applications was lead
  - CBDRA Staff – Sudip Bhattacharjee, Ann Wittenauer, John Cuffe
  - augmented by Joe Staudt in CES
- Currently, CES is taking the reigns from ADRM and CBDRA

FUTURE ON
Activating Change.

The LBD Project and IMI

- The project at the Census Bureau that covers all IMI work is "Research and Development to Enhance the Longitudinal Business Database and Derivative Products", which is referred to as the "Business Dynamics Statistics for Science-Based Innovation (BDS-SBI"

- This project is wide in scope, but covers work that started under:
  - Project 422 – Conduct Research on Longitudinal Business Database
  - Project 913 - Pilot Project to Integrate New External Data with the Integrated Longitudinal Database (ILBD)

FUTURE ON
Activating Change.

Infrastructure that IMI builds upon…

- Longitudinal Employer Household Dynamics (LEHD) Program
  - Began in the very late 1990's
  - Supported with a budget initiative about a decade later
- Longitudinal Business Database (LBD)
  - Began in early 2000's
  - Budget initiative to enhance in FY2015
- "Modernizing Economic Statistics" Initiative
  - Requested FY2017, not funded
  - Included big data research efforts more generally; IMI-related activities were an initial proof of concept for data ingest.

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

8

FUTURE ON
Activating Change.

Special Considerations...

- LEHD – no state-level estimates (without state approval, three state rule)
  - Snapshot (2014 just released, see working paper CES-WP-2018-27)
  - Snapshot is being baked into the production code to facilitate more frequent updates in the future. This will probably be rolled out over time for the various files.
  - UI coverage for some university employees may be problematic (e.g. graduate students)
- Implicit Samples (as more universities are being added by IRIS)
- Disclosure Review Board (DRB)
  - New rules being implemented
  - Sub-state estimates $\rightarrow$ noise infusion
  - Details (e.g. how much noise to add) are still being worked out

Questions?

- There will be time for some questions at the end...
- For additional information please contact:
  - Shawn.D.Klimek@census.gov

IRIS Overview

Jason Owen-Smith

Executive Director

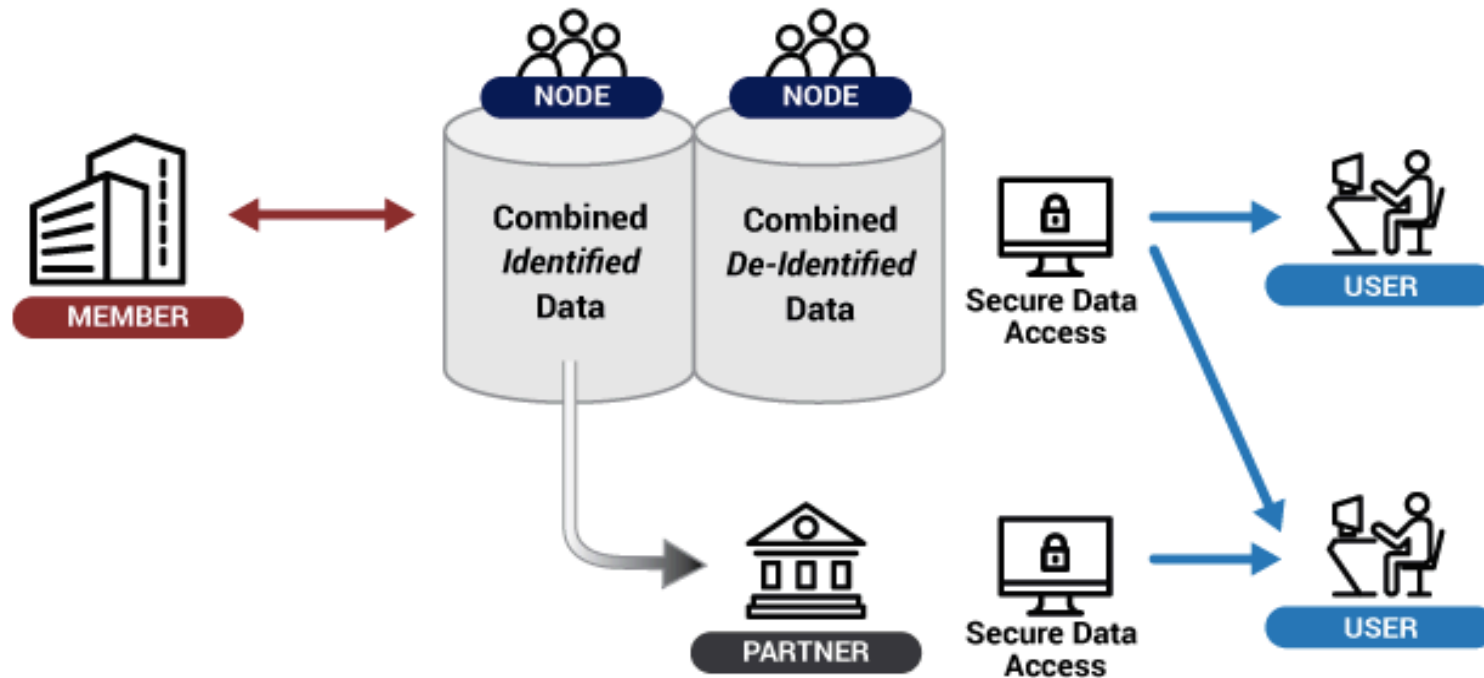Institute for Research on Innovation & Science (IRIS)

IRIS Overview

- Intermediary organization to facilitate data sharing for research and reporting among universities, federal agencies, researchers
    - 33 university members -- ~30% of 2016 federal R&D expenditures
    - Broad engagement and support from research community/associations

- Platform for fundamental social science research of immediate practical import
    - 60 researchers from 19 institutions have accessed research data to date
    - Sloan support for research community building (new grant call soon)

- One of the first research/training infrastructures for computational social science
    - Training and courses with ICPSR, JSMP

MEMBERS: Universities contribute data, support infrastructure and receive campus-specific and aggregate reports

NODES: Approved nodes materially improve data, develop products, and expand user communities

USERS: Approved users securely access de-identified aggregate datasets

NODE

NODE

MEMBER

Combined *Identified* Data

Combined *De-Identified* Data

Secure Data Access

USER

PARTNER

Secure Data Access

USER

PARTNERS: Approved partners receive data from IRIS which they improve and make accessible throught their own secure systems

©2015 IRIS

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

13

FUTURE ON
Activating Change.

Second Research Data Release

- 26 universities
  - $61.6 Billion Direct Cost Expenditures (+70% from last release)
- Transaction level data
  - 296,253 federal and non-federal sponsored projects (+67% from last release)
    - 82,028 match NIH/NSF/USDA grants
  - 478,815 individuals (+43% from last release)
  - $18.5B in vendor spending to 441,796 establishments
  - $8.5B in subcontracts to other 13,262 performers
- Links to abstracts etc for federal awards (NIH, NSF, USDA)
- Individual level links to dissertation information, patent data
- Title 13 crosswalks to LEHD, LBD, ACS, Decennial Census (available only through the FSRDC system)

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

14

FUTURE ON
Activating Change.

Accessing data through the IRIS VDE

- No Census data available in any form, but IRIS data is mirrored in RDCs
- Windows virtual desktop environment, shuts down i/o on your machine.
- Data can be added to your scratch space by IRIS research support staff
- Only aggregate information and statistics such as regression coefficients can be removed
- Export occurs after a privacy disclosure process based on Census procedures
- Restricted access documentation in Wiki format allows user updates
- Online ticket system to report data and software bugs
- Researchers who are not affiliated with an IRIS member institution must pay a modest seat fee to access data

Accessing data through the IRIS VDE

- Check out background materials and FAQ on IRIS Website (http://iris.isr.umich.edu/research-data/)

- Download and complete application and data use agreement

- All virtual machines are loaded with Windows 7 and the following software, packages, and libraries: Microsoft Office, Stata 14, SAS 9.4, R / RStudio, MATLAB, LaTeX, HeidiSQL, MS SQL Server Management Studio 2014, Gephi, Cytoscape, QGIS, GRASS GIS, Adobe Acrobat Pro,, Notepad++, Python, Anaconda, Jointpoint, PuTTY, WinSCP, and TightVNC Viewer. Researchers can contact IRIS for any questions concerning existing software or to request the installation of additional applications in the VDI.

Research Community Development

- More than **60 researchers** from **19 institutions** have accessed data through either VDE or FSRDC so far

- First research meeting this summer

- Sloan Foundation Support for research grants
  - $15 k Dissertation
  - $30 k early and mid-career grants
  - Call for proposals in Fall 2018

- Constituting a scientific advisory board this summer

- Next data release (target=35 universities) in Winter, 2019.

Call for Proposals: 2019 IRIS Research Awards

- Up to $15,000 for dissertation awards and up to $30,000 for early career and established researcher awards will be disbursed to the recipient's institution. Funds can be used for personnel, equipment, supplies, travel, and other expenses.

- Awards must include no more than 12% overhead or indirect costs to be paid as a part of the award total. Applicants with proposals for projects with longer timelines may be considered, but the monetary award is limited to a 15-month time period beginning January 2019 through March 2020.

- For details see iris.isr.umich.edu/research-data/grants/

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

18

FUTURE ON
Activating Change.

Questions?

- There will be time for some questions at the conclusion of the webinar.
- For additional information about IRIS please contact:
  - jdos@umich.edu or nbirk@umich.edu
- Follow us on Twitter! @IRIS_UMETRICS #irisumetrics

Sign up for updates on data improvements, funding and training opportunities as well as IRIS news and events via our website's contact page http://iris.isr.umich.edu/contact/

Using Census Data at the Federal Statistical Research Data Centers

Barbara Downs

Director, FSRDC

Center for Economic Studies

U.S. Census Bureau

Agenda

- FSRDC Environment
  - Physical Environment
  - Collaboration
  - Disclosure Avoidance Review
- Project Proposals
- Census Data at FSRDCs

FSRDC Environment
Physical Security

- Secure Census facility within host institution
- Census employee on-site at all FSRDCs
- Authorized personnel only
- Researcher Special Sworn Status
  - Requires moderate level background check
  - Oath of confidentiality is for life
- Data accessed via secure connection from thin client device to Census data facility in Bowie, MD
- Printing strictly controlled
- No internet access

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

22

FUTURE ON
Activating Change.

FSRDC Environment
Collaboration

- Each project has "home" FSRDC

  - Researchers may collaborate across FSRDCs

  - Projects may move "homes" as researchers relocate

- FSRDC Administrator

  - Coordinates project access across FSRDCs

  - Coordinates review of output

FUTURE ON
Activating Change.

FSRDC Environment
Disclosure Avoidance Review

- Analyses are model-based; tabulations discouraged
- Prepare disclosure request with FSRDC Administrator assistance
- FSRDC Administrator reviews
- Center for Disclosure Avoidance Research reviews
- Output released *electronically* after approval
- Researcher papers
  - Census does not review researcher papers
  - Papers should include standard Census disclaimer

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

24

FUTURE ON
Activating Change.

Census Project Proposals

- Required for all projects using Census Bureau data
- Three stages of review
  - FSRDC Development and Review
    - Abstract
    - Proposal
    - Predominant Purpose Statement
  - Census Bureau Review
    - 5 concurrent reviews
  - Other Agency Review
    - SSA, BLS, IRS – any agency providing some of the project's data

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

25

FUTURE ON
Activating Change.

Census Project Proposals
Review Criteria

- Scientific merit

- Requires non-public data

- Provides benefit to Census Bureau programs

- Is feasible

- Poses no risk of disclosure of individual or business

Census Project Proposals
Benefits to Census Bureau

- Census-IRS Criteria Agreement

- Helps Census check data it collects, edits, and tabulates

  - Permits rigorous analysis of confidential data

  - Tests validity of data processing rules

  - Evaluates conceptual and processing assumptions

- Prepares new economic or population estimates

Census Project Proposals
Timing

- Census review
  - ~75 days
- Other Agency review
  - 3 to 6 months
- Special Sworn Status
  - Concurrent with Agency review
  - 3 to 5 months

Census Data
Economic

- Economic data
  - Collected at establishment and/or firm level, admin data is at the Employer Identification Number (EIN)
  - Detailed geography (latitude/longitude for later years, address for all)
- Linking or Matching data
  - Across censuses or surveys
  - Over time
  - Across entities (establishment to firm)
  - To external data (e.g., Compustat, patent data) via business name and address matching or EIN

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

FUTURE ON
Activating Change.

Census Data
Popular Economic Data

- Longitudinal Business Database (LBD)
  - All US establishments with at least 1 employee 1976-2012
  - Identifies establishment births, deaths, and continuers
- Census and Annual Survey of Manufactures (CMS, ASM)
  - Detailed information about manufacturing establishments
  - Measures costs and revenues, calculate productivity
  - Detailed materials and product information
- Economic Censuses (EC)
  - 1977-2012, every 5 years
  - Covers most industries at detailed levels
  - Industry-specific questions on sporadic basis
- Integrated Longitudinal Business Database (iLBD)
  - Merges non-employer with employer (LBD) data
- Linked Foreign Trade Transactions Database (LFTTD)
- Longitudinal Employer-Household Dynamics (LEHD)
  - Link UI and QCEW data with Census demographic and economic data

Census Data
Longitudinal Employer-Household Dynamics

- Underlying data from states, SSA, and Census Bureau

- Use requires provision of benefit to LEHD program

- State-level review of project proposals
  - Some states provide global approval

Census Data
Demographic

- More details than Public Use Files
  - No top-coding
  - Uncollapsed variables
  - Finer geographic detail
- Match to external sources
  - Other public use or macro-level data
  - Link by geography or PIKs

Census Data
Popular Demographic Data

- Decennial Census (1950-2010)
  - Short and Long Form data through 2000 (not sample)
- American Community Survey
- Survey of Income and Program Participation
- Current Population Survey
  - Annual Social and Economic Supplement (March Supplement)
  - Food Security (December Supplement)
- American Housing Survey

FUTURE ON
Activating Change.

More Information

- Links
  - Federal Statistical Research Data Centers
    www.census.gov/fsrdc
  - National Center for Health Statistics
    www.cdc.gov/nchs/r&d/rdc.htm
  - Agency for Healthcare Research and Quality
    www.meps.ahrq.gov/mepsweb/data_stats/onsite_datacenter.isp
  - Bureau of Labor Statistics
    www.bls.gov/bls/blsresda.htm
- Contact
  - Barbara Downs (barbara.a.downs@census.gov)

Questions?

Thank you!

For a recording or transcript of this webinar contact:

Gilda Beauzile

Center for Economic Studies

Gilda.Beauzile@census.gov