Linked Employer-Employee Data for the U.S.: Overview of the LEHD program

Statistics Canada Socio-Economic Workshop
February 26, 2018

Erika McEntarfer, Lead Economist
Longitudinal Employer-Household Dynamics (LEHD) Program
Center for Economic Studies, U.S. Census Bureau

Disclaimer: All data examples shown are from public use data or are fictitious examples of firm or worker data for demonstration purposes only. All opinions are my own, not those of the U.S. Census Bureau

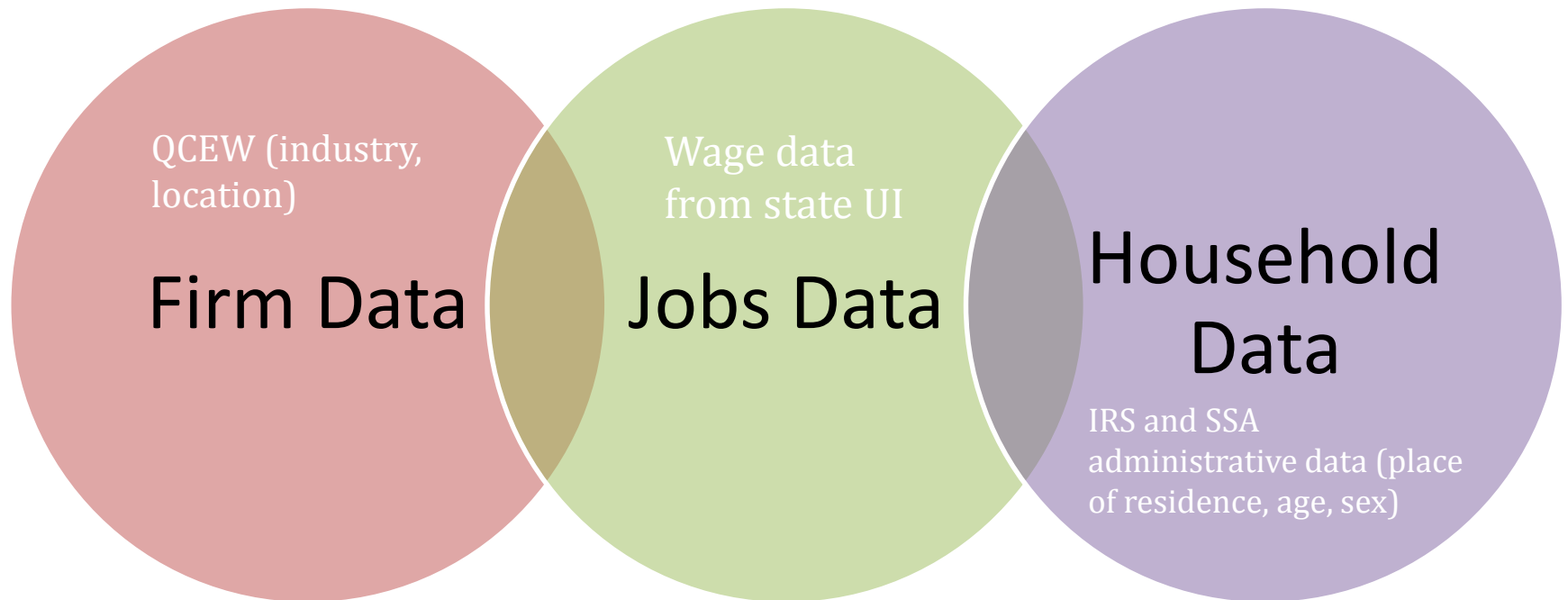# What is the Longitudinal Employer Household Dynamics program?

An innovative federal statistical program, collecting existing data and ...
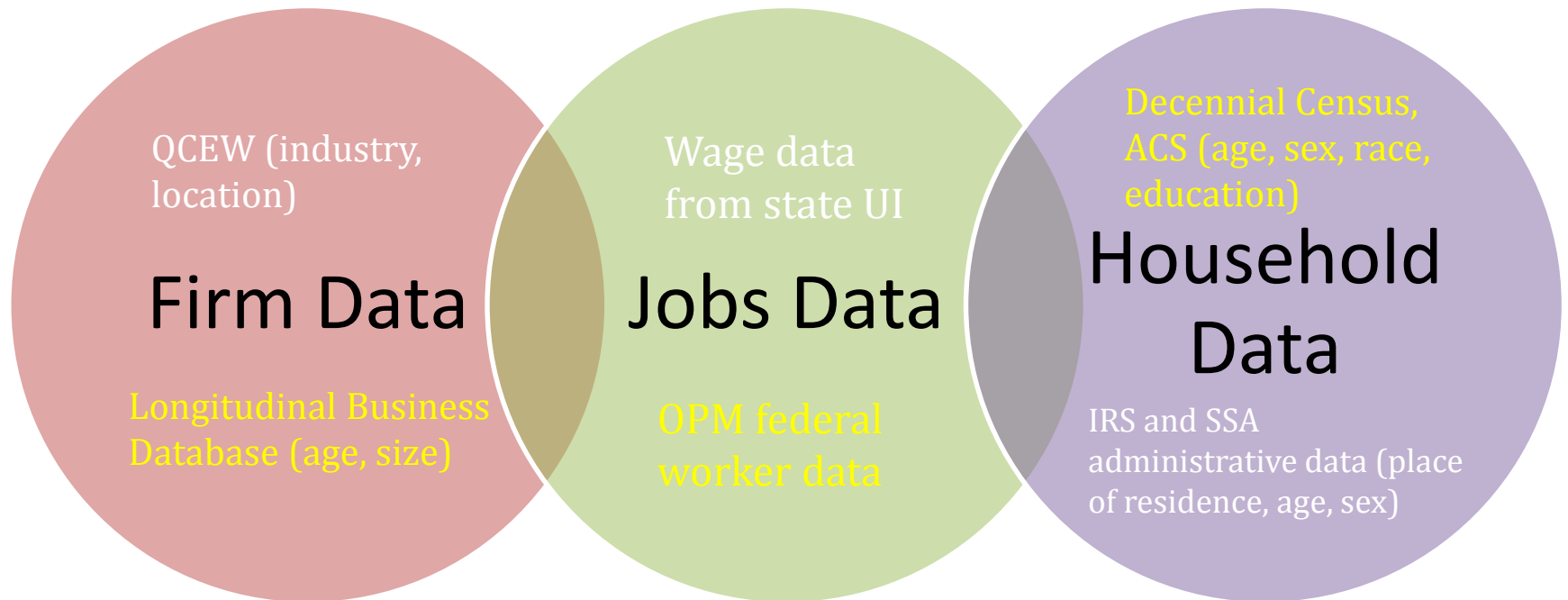
Firm Data

Jobs Data

Person Data
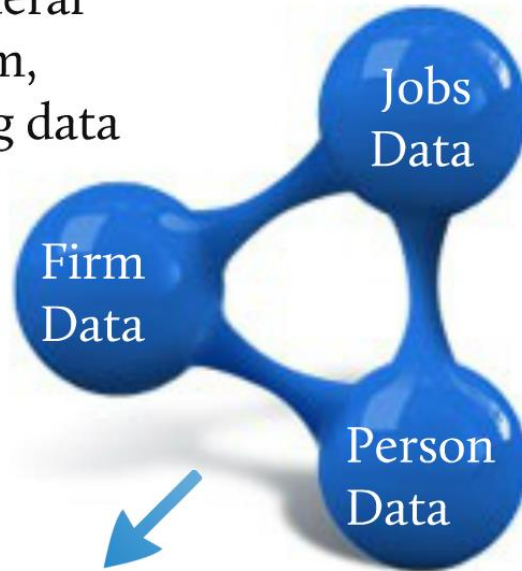
..linking it together to provide new information sources at low cost

New linked national jobs data for the U.S.

The LEHD linked microdata   circa 2008

Firm Data — QCEW (industry, location)

Jobs Data — Wage data from state UI

Household Data — IRS and SSA administrative data (place of residence, age, sex)

The LEHD linked microdata   circa 2018

Firm Data

QCEW (industry, location)

Longitudinal Business Database (age, size)

Jobs Data

Wage data from state UI

OPM federal worker data

Household Data

Decennial Census, ACS (age, sex, race, education)

IRS and SSA administrative data (place of residence, age, sex)

All 50 states and DC, 1990-2017, time-series varies by state, covers over 95% of U.S. employment

Overview of talk today:

LEHD program
- Inception and purpose
- Lessons learned

Accessing LEHD data
- Public use data
- Federal Research Data Centers (restricted microdata access)

The LEHD program: a brief history

1999   LEHD starts as as an academic collaboration with three academic economists John Abowd, John Haltiwanger, and Julia Lane

2000   First two states (IL and MD) share data with Census

2003   First public use data product QWI released

2006   Second public use data product LODES released

2009   LEHD budget initiative insures future of program and products

2014   Third public use data product J2J released, national QWI released

2018   New Post-Secondary Earnings Outcomes pilot released

Lessons learned

Goals of the original LEHD pilot:

1. Create linked employer-employee data for the U.S.
   - Establish partnerships to share data
   - Develop methodologies to link, edit, and complete the data

2. Demonstrate the utility of linked employer-employee data
   - for researchers
   - as a resource for the U.S. federal statistical system, linking person and business data frames

United States™ Census Bureau

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

Creating linked employer-employee data for the US:

Initial challenge: best sources of jobs data reside outside US statistical agencies

Solution: partner with agencies that collect jobs data to administer programs

- 49 states & DC share data with Census
- These relationships are voluntary
- In return for data-sharing, new statistical products

Lessons learned: Partnerships

Voluntary data sharing partnerships are both enormously valuable and difficult to build and maintain

Benefits
- Allows valuable data sharing at minimal cost
- Partnerships insure shared benefits to both parties
- Mutual benefit helps insure higher quality data provided

Challenges
- Many potential benefits of multi-linked data systems do not directly accrue to the data sharing agency, yet they still incur risks (disclosure, policy risks)
- Voluntary partnerships can affect timeliness and dependability of available data

Lessons learned

Goals of the original LEHD pilot:

1. Create the linked employer-employee data
   - Establish partnerships to share data
   - Develop methodologies to link, edit, and complete the data

2. Demonstrate the utility of linked employer-employee data
   - for researchers
   - as a resource for the U.S. federal statistical system, linking person and business data frames

Lessons learned: Linked data systems

Linked data systems have benefits and unique challenges

Benefits
- Generates new information at minimal cost, allows better leveraging of data to handle non-response

Challenges
- Complexity of the data system
- Understanding and measuring data quality
- Data generating process of externally provided data is external to agency
- Confidentiality protection is more complex when statistical agency does not control access to all microdata

United States™ Census Bureau

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

Lessons learned

Goals of the original LEHD pilot:

1. Create linked employer-employee data for the U.S.
    - Establish partnerships to share data
    - Develop methodologies to link, edit, and complete the data

2. Demonstrate the utility of linked employer-employee data
    - for researchers
    - as a resource for the U.S. federal statistical system, linking person and business data frames

Lessons learned: Demonstrating utility

Researchers:

- Always understood the utility of linked employer-employee data

- Challenge is explaining the complexity of the system and the legal/policy environment

- Researchers can get frustrated when many different data owners have to sign off on projects

Lessons learned: Demonstrating utility

"The costs of statistical data collection for surveys and censuses are increasing faster than inflation; the demand for more timely and smaller area statistics is increasing; ...more and more administrative data are digitized ... and the federal government budget for statistics is likely to be flat or declining in the mid-term....The future most of us feel we are building is one of multiple sources of data ...being used to support one another."

-- U.S. Census Bureau Director Robert Groves interview to the American Statistical Association in 2012

Lessons learned: Demonstrating utility

Statistical agencies

- Potential of administrative data for meeting current challenges now well understood

- But much work remains to build the infrastructure and human capital necessary to better leverage administrative data in US federal statistics

- These changes will require investments in new technologies and mixed-data source methodologies

Other lessons

The benefits of academic/stat agency collaborations:

- Academic partners bring new ideas and methods
- Statistical agency staff insure data quality and provide user support for new products
- Academia -> stat agency -> users pipeline works best if agency culture and human capital supports these collaborations.

Creating a culture that can innovate:

- "Culture eats strategy for breakfast" – Peter Drucker
- Need to attract different kinds of talent to meet current challenges, need to create a culture that attracts them and empower them to find and implement solutions

Overview of talk today:

LEHD program
- Inception and purpose
- Lessons learned

Accessing LEHD data
- Public use data
- Federal Research Data Centers (restricted microdata access)

United States Census Bureau

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

Spectrum of LEHD Data Access

1. Quarterly Workforce Indicators,
2. LEHD Origin-Destination Data,
3. Job-to-Job Flows

Census staff access LEHD microdata (no PII) under approved Census projects

3. Public Use Files

6. Census internal projects (LEHD Infrastructure, LBD)

External researchers can access confidential microdata (no PII) under approved Census projects. All output reviewed for disclosure before can leave the center

5. Access at Data Use Center or Enclave

Open - *Restricted*

Spectrum of LEHD Data Access

1. Quarterly Workforce Indicators,
2. LEHD Origin-Destination Data,
3. Job-to-Job Flows

3. Public Use Files

Open - *Restricted*

The Quarterly Workforce Indicators:

Key data elements:
- Employment
- Job Creation/Destruction
- Hires/Separations
- Earnings (new hires, attached workers)

Frequency:
- Quarterly

Timeframe:
- Varies by state, beginning in early 1990s for some, late as 2010 for MA

Level of industry/geographic/demographic detail
- 4-digit NAICS
- County-level geography for most statistics
- By characteristics of workforce: age, sex, education, race/ethnicity

United States™ Census Bureau | U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

20

QWI: examining workforce demographics

- This graph is from a Bloomberg article about recent labor shortages in construction

- **Popular uses**:
  - aging workforce
  - local labor market dynamics
  - detailed local industry analysis

Bloomberg ▼ | Millennials: Builders Are Desperate to Hire You

Graying Hard Hats
Share of industry workers by age, based on a moving average of seasonally-adjusted data

■ Construction, 19-24   ■ All industries, 19-24   ■ Construction, 45-54   All industries, 45-54
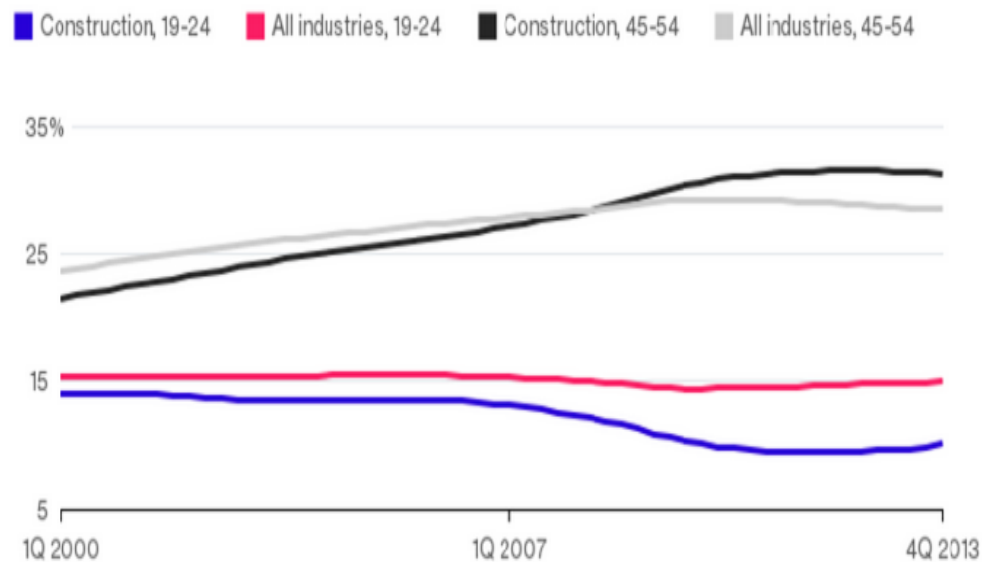
Source: U.S. Census/Hubert Janicki and Erika McEntarfer

Bloomberg

United States Census Bureau

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

LEHD Origin-Destination Employment Estimates (LODES):

- **Key data elements:**
  - Employment
  - Residence
- **Frequency:**
  - Annual
- **Timeframe:**
  - Varies by state,
- **Level of industry/geographic/demographic detail**
  - Block-level employment and residence
  - By industry sector, age and size of firm
  - By characteristics of workforce: age, sex, education, race/ethnicity, average earnings

LODES: Where workers live, and where they work

- This map shows LODES data of where residents of Vancouver, Washington work

- **Popular uses**:
  - local economic development
 - business site selection
  - emergency planning

Legend:
- 5 - 165 Jobs/Sq.Mile
- 166 - 648 Jobs/Sq.Mile
- 649 - 1,452 Jobs/Sq.Mile
- 1,453 - 2,578 Jobs/Sq.Mile
- 2,579 - 4,025 Jobs/Sq.Mile

OnTheMap

United States™
Census
Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov
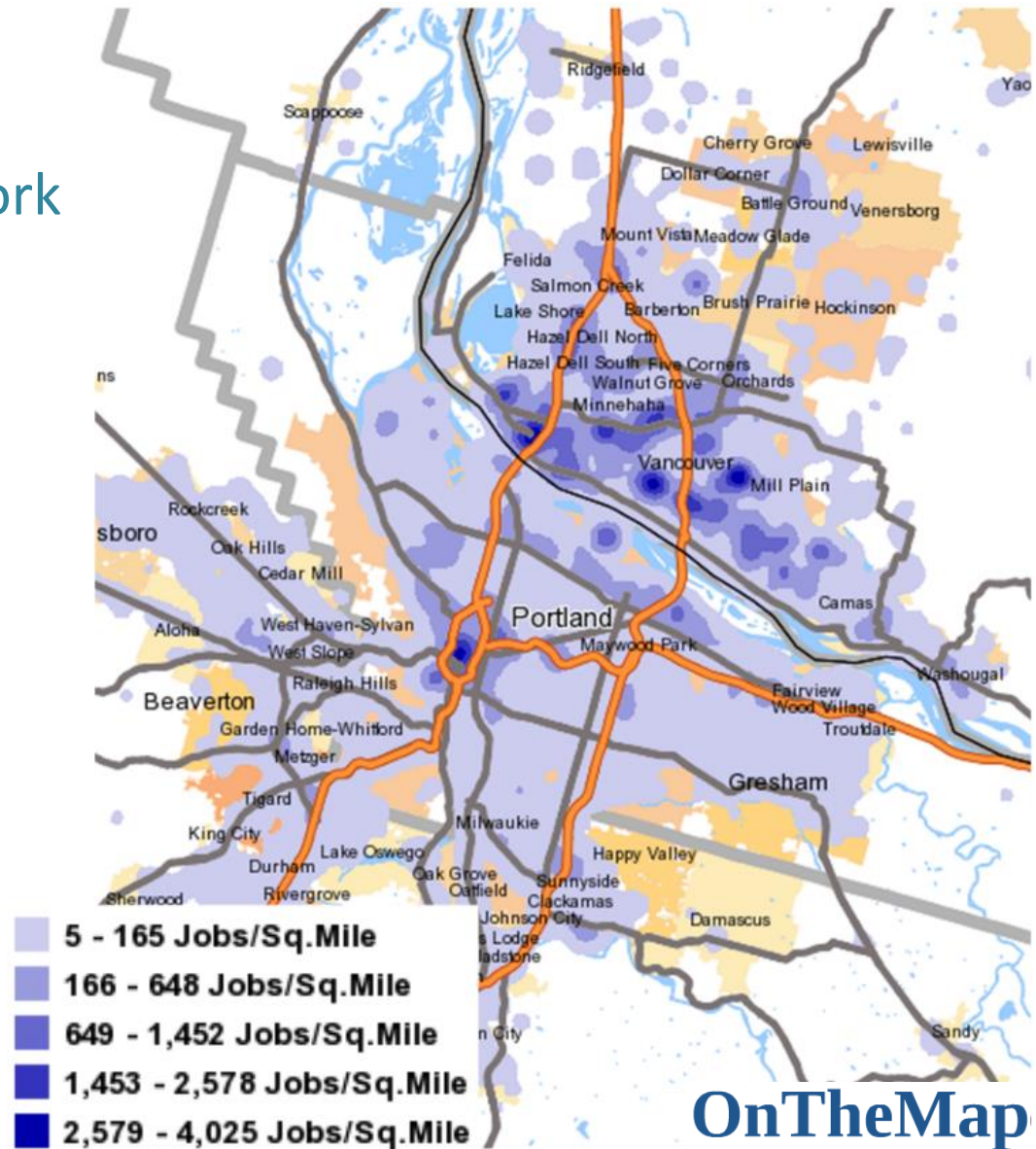
Job-to-Job Flows (J2J):

- **Key data elements:**
  - Job-to-Job transitions
  - Employment/nonemployment transitions
- **Frequency:**
  - Quarterly
- **Timeframe:**
  - Varies by state, national data starts in 2000
- **Level of industry/geographic/demographic detail**
  - State, Metro (CBSA)
  - By industry sector, age and size of firm
  - By characteristics of workforce: age, sex, education, race/ethnicity, average earnings

J2J: A closer look at labor market dynamics

- These graphs show the shale oil boom and bust in North Dakota

- **Popular uses**:
    - Understanding worker reallocation across industries and geographies

North Dakota Mining Sector, 2001.2-2016.Q3
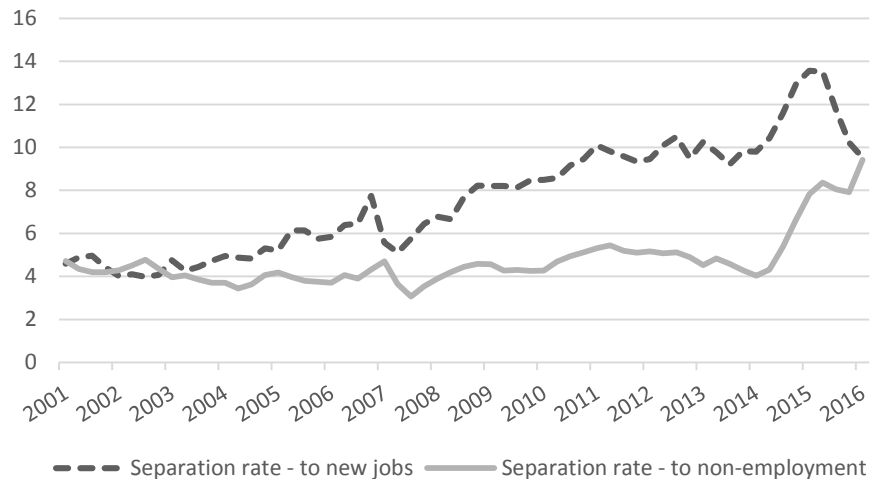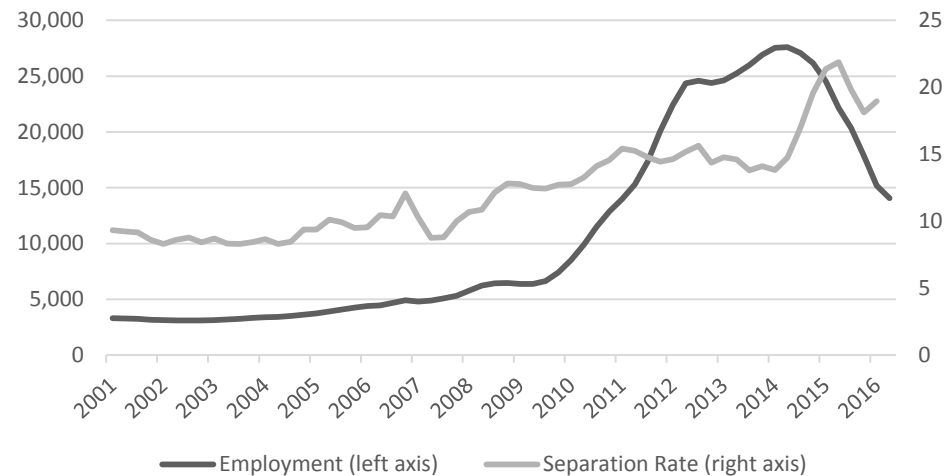
Note: Calculated from Census Job-to-Job Flows data, 2017Q3 release
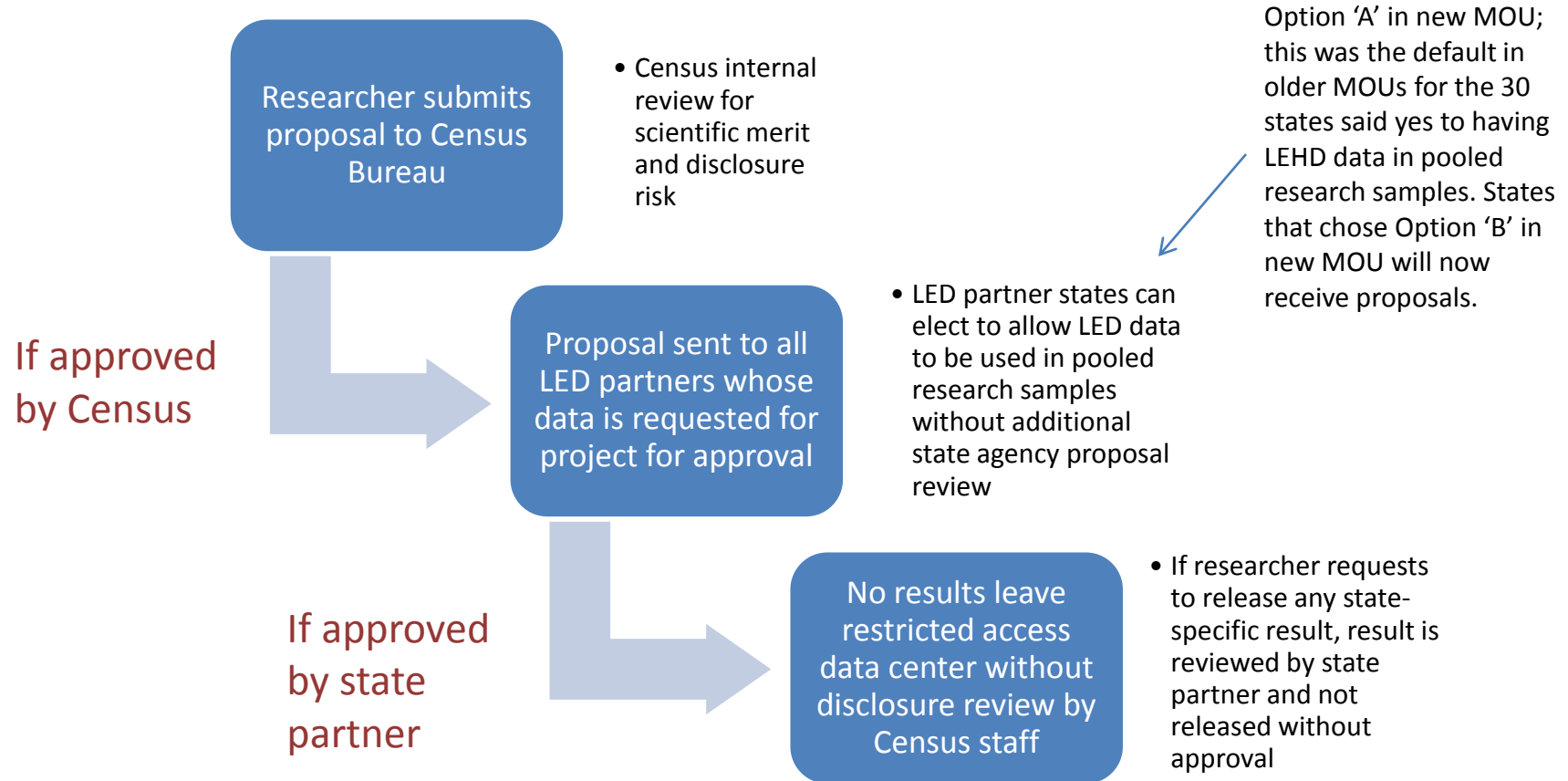
Spectrum of LEHD Data Access

External researchers can access confidential microdata (no PII) under approved Census projects. All output reviewed for disclosure before can leave the center

5. Access at Data Use Center or Enclave

Open - *Restricted*

Approval process for external researcher access to LEHD microdata

Option 'A' in new MOU; this was the default in older MOUs for the 30 states said yes to having LEHD data in pooled research samples. States that chose Option 'B' in new MOU will now receive proposals.

Researcher submits proposal to Census Bureau

- Census internal review for scientific merit and disclosure risk

If approved by Census

Proposal sent to all LED partners whose data is requested for project for approval

- LED partner states can elect to allow LED data to be used in pooled research samples without additional state agency proposal review

If approved by state partner

No results leave restricted access data center without disclosure review by Census staff

- If researcher requests to release any state-specific result, result is reviewed by state partner and not released without approval

To sum up:

- The LEHD program at the U.S. Census Bureau creates linked employer-employee data for the U.S.
  - 3 public use data products
  - Microdata access in Federal Research Data Centers
- In the U.S., this was accomplished by working with multiple federal and state partners (not necessary in every country)
- Resulting complexity of the system provides useful lessons for more extensive use of administrative data in federal statistical agencies.

Extra Slides

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

The LEHD linked microdata

Employer Characteristics File (ECF)

Employment History File (EHF)

Individual Characteristics File (ICF)
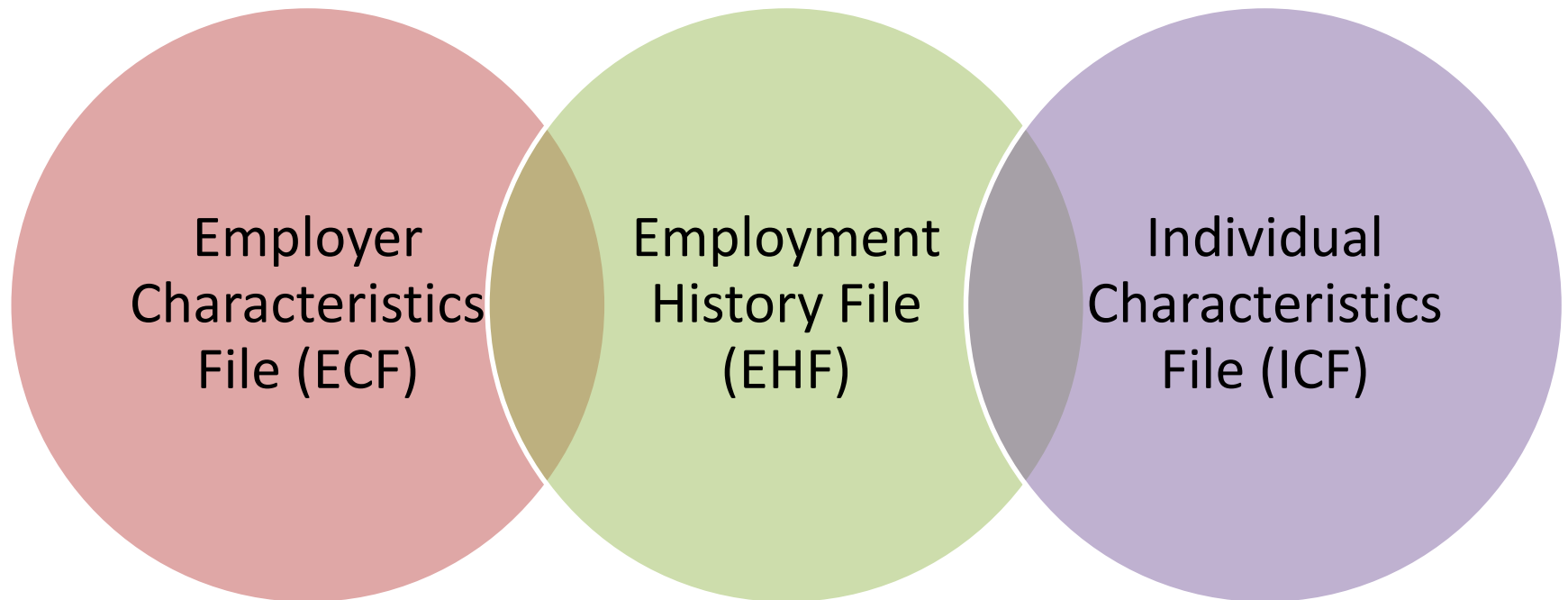
Except for the ICF, files are organized by state. ICF files are national-level files.

The EHF: a longitudinal quarterly job-history file

Changes jobs in Q3

The Employment History File (EHF)

PIK	SEIN	Q1	Q2	Q3	Q4	Q5
Person1	Firm A	7000	7000	3000	0	0
Person1	Firm B	0	0	4000	8000	8000
Person2	Firm A	500	0	0	0	0
Person2	Firm D	0	1000	1000	0	0
Person2	Firm F	0	0	3000	4000	4000

Unit of observation is a job

Source data: quarterly unemployment insurance (UI) wages from states, federal worker data from Office of Personnel Management

The ECF: a quarterly firm characteristics file

Source data: Quarterly Census of Employment and Wages, Longitudinal Business Database

The Employer Characteristics File (ECF)

SEIN	SEINUNIT	Qtr	Industry	Emp	Firmsize	Firmage
FirmA	Unit1	1	333333	302	5000	12
FirmA	Unit2	1	666111	4030	5000	33
FirmA	Unit3	1	444222	20	5000	21
FirmB	Singleunit	1	771111	1	1	2
FirmC	Singleunit	1	666622	5	7	24

Unit of observation is a State UI taxpayer ID

The ICF: a person-level file file

The Individual Characteristics File (ICF)

PIK	DOB	Sex	Race
Person1	MM/DD/YYYY	M/F	Race1
Person2	MM/DD/YYYY	M/F	Race4
Person3	MM/DD/YYYY	M/F	Race1

Source data:
 - *demographics*: Social Security transaction file (Census Numident), Decennial Censuses, American Community Survey
 - *residential address*: programs data, federal tax returns

Linking the data

ICF

PIK	DOB	Sex	Race
Person1	1/3/73	F	White
Person2	3/1/37	F	Asian

EHF

PIK	SEIN	Q1	Q2	Q3
Person1	Firm A	7000	7000	3000
Person1	Firm B	0	0	4000
Person2	Firm A	500	0	0
Person2	Firm D	0	1000	1000
Person2	Firm F	0	0	3000

ECF

SEIN	SEINUNIT	Qtr	Industry	M1emp	M2emp	M3emp
FirmA	Unit1	1	333333	302	335	330
FirmA	Unit2	1	666111	4030	4032	4031
FirmA	Unit3	1	444222	20	23	21

Where to find more information on LEHD