

INFO 7470/ECON 7400/ILRLE 7400  
Universes, Populations, Frames,  
and Sampling

John M. Abowd and Lars Vilhuber  
February 4, 2013  
and  
February 15, 2016

# Outline

- A little more statistical infrastructure
- What is a census?
- Basic definitions
- Demographic sampling frames
- Economic sampling frames
- Coverage and un-duplication
- Basic relations connecting frames

# Principal Statistical Agencies

- Office of Management and Budget
- Bureau of Economic Analysis (Commerce)
- Bureau of Justice Statistics (Justice)
- Bureau of Labor Statistics (Labor)
- Bureau of Transportation Statistics (Transportation)
- Census Bureau (Commerce)
- Economic Research Service (Agriculture)
- Energy Information Administration (Energy)
- Environmental Protection Agency (Independent)
- Internal Revenue Service, Statistics of Income (Treasury)
- National Agricultural Statistical Service (Agriculture)
- National Center for Education Statistics (Education)
- National Center for Health Statistics (Centers for Disease Control and Prevention, HHS)
- National Science Foundation, Science Resources Statistics (Independent)
- Social Security Administration, Office of Policy (Independent)

# Role of OMB

- The Office of Management and Budget oversees the regulatory and budgeting environment
- Home of the Chief Statistician of the United States (currently Katherine Wallman)
- Budgets for all of the statistical agencies are reviewed by OMB
- Regulations for reporting of Race, Ethnicity, Industry, Geography

# Census

- An attempt to enumerate every entity in a specified universe (target population)

“Perhaps the earliest type of survey is the census, generally conducted by governments. Censuses are systematic efforts to count an entire population, often for purposes of taxation or political representation.” (Groves *et al. Survey Methodology*, 2004, pp. 3-4.)

# Legal v. Statistical Concepts

- In the United States, the decennial census of population must be conducted as an “enumeration”
- Department of Commerce v. U.S. House of Representatives (1999) defined enumeration so that “The Census Act prohibits the proposed uses of statistical sampling to determine the population for congressional apportionment purposes.”

# Legal v. Statistical Concepts

- But Utah v. Evans (2002) allowed the use of “hot-deck imputation” to allocate individuals to vacant domiciles deemed inhabited on April 1, 2000. “Indeed, the Bureau's imputation method is similar in principle to other efforts used since 1800 to determine the number of missing persons, including asking heads of households, neighbors, landlords, postal workers, or other proxies about the number of inhabitants in a particular place.”

# Science of the Controversy

- Assessing the accuracy of a census requires supplemental information
- The supplemental information is usually collected in the form of a coverage assessment survey
- Many statistical methods are used to compare the original census to the post-enumeration survey
- The assumptions inherent in these analyses are often difficult to test



# Survey

- “... systematic method for gathering information from (a sample of) entities for the purposes of constructing quantitative descriptors of the attributes of the larger population of which the entities are members” (Groves *et al.*, 2004, p. 2)
- Sampling is optional in general, but in this course we will use “sample survey” and “survey” interchangeably, reserving “census” for any enumeration activity
- But, entities covered by the survey represent entities in the population

# Censuses v. Surveys

- Decennial Census of Population and Housing (short form until 2010)
- The Economic Censuses (years ending in 2 and 7)
- Quarterly Census of Employment and Wages
- Long form (through 2000)
- American Community Survey (since 2005)
- Annual Survey of Manufactures, Monthly Survey of Construction
- Current Employment Statistics

# Administrative Record

- Any information concerning specific entities in a designated population that is collected by a governmental agency for the purposes of enforcing a specific law
- The legal distinction between administrative record and statistical information systems was made in the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)

# CIPSEA

- For the Census Bureau, these protections are part of Title 13
- Many other agencies (*e.g.*, BLS) had no statutory statistical agency status until CIPSEA
- Statistical use of administrative records under CIPSEA prohibits re-use of the statistically-enhanced records for the legal enforcement activities even when the statistical and administrative functions are performed inside the same agency
- Sharing of some administrative records for statistical purposes is explicitly authorized among agencies previously not authorized to do so (Census, BEA, BLS, and IRS)
- Other sharing may occur when it is not explicitly prohibited
- Dense legal forest that we will hike later

# Administrative v. Statistical Uses

- Tax information reporting to the tax collecting agency from a specific taxable entity (IRS)
- Determination of Unemployment Insurance benefits and eligibility from employer wage reports (state UI offices)
- Tables of household Adjusted Gross Income by geography and size (Statistics of Income Division)
- Quarterly job creations and destructions based on establishment reports (BLS-BED)
- Quarterly job creations and destructions based on job-level reports (Census-QWI)

# Basic Definitions

- Universe (target population)
- Out-of-scope population
- Frame population
- Geography
- Population demographics
- Business entity demographics
- Government entities

# Universe = Target Population

- Theoretical construct specifying every entity that satisfies a set of explicit qualifying conditions
- In probability models describing the statistical process of estimation (either finite-population or super-population methods), the universe is the event that occurs with probability 1

# Target Population

- The “target population” is any entity satisfying the set of conditions that specify the universe. “Universe” and “Target Population” are synonyms
- Example 1: “Human population” All people, male and female, child and adult, living in a given geographic area at a particular date
- Example 2: “Establishment population” A business, industrial, service or governmental unit at a single location that distributes goods or performs services on a particular date (or during a given period)



# Out-of-scope Population

- An entity that is outside either the geographic region under study or fails to meet another specific restriction imposed on the target population
- Example 1: When the in-scope population is “persons age 16 or over living in households,” persons age 15 or younger and all persons living in group quarters are out of scope
- Example 2: When the in-scope population is “employer establishments,” all establishments with zero employees are out of scope

# Frame Population

- Set of target population, or universe, entities that can be selected into a sample or census
- Also called a sampling frame
- The frame population or sampling frame is the physical manifestation of the universe—if an entity is not on the frame (or one of the frames for multi-frame sampling), then it cannot be in the census or survey
- Simple cases: (single frame sampling) a list of all addresses to be sampled; list of all people to be sampled; list of all businesses to be sampled
- Complex sample designs use multiple frame populations to get better coverage of the target population or universe
- Complex frame example: CPS or SIPP

# Complex Housing Frame Example

- Survey of Income and Program Participation (SIPP)
  - Similar complex frame used for CPS
- Multi-stage sample
- Primary Sampling Units are geographic areas
- Within PSUs five distinct frames are used
- [SIPP Sample Design](#)

# SIPP Details I

- 5 Frame populations: Unit, Area, Group Quarters, Coverage Improvement, and New Construction.
- Unit, area, and group quarters frames are based on census counts from the most recent decennial census of population. These account for 90% of the sample addresses
  - In the unit and group quarters frames, the clusters contain only a single housing unit or housing unit equivalent
  - Within each of these frames, clusters of housing units are selected
  - In the area frame, the cluster contains four “expected” housing units
- The coverage improvement frame includes housing units missing from the most recent decennial census but found in the post-enumeration surveys

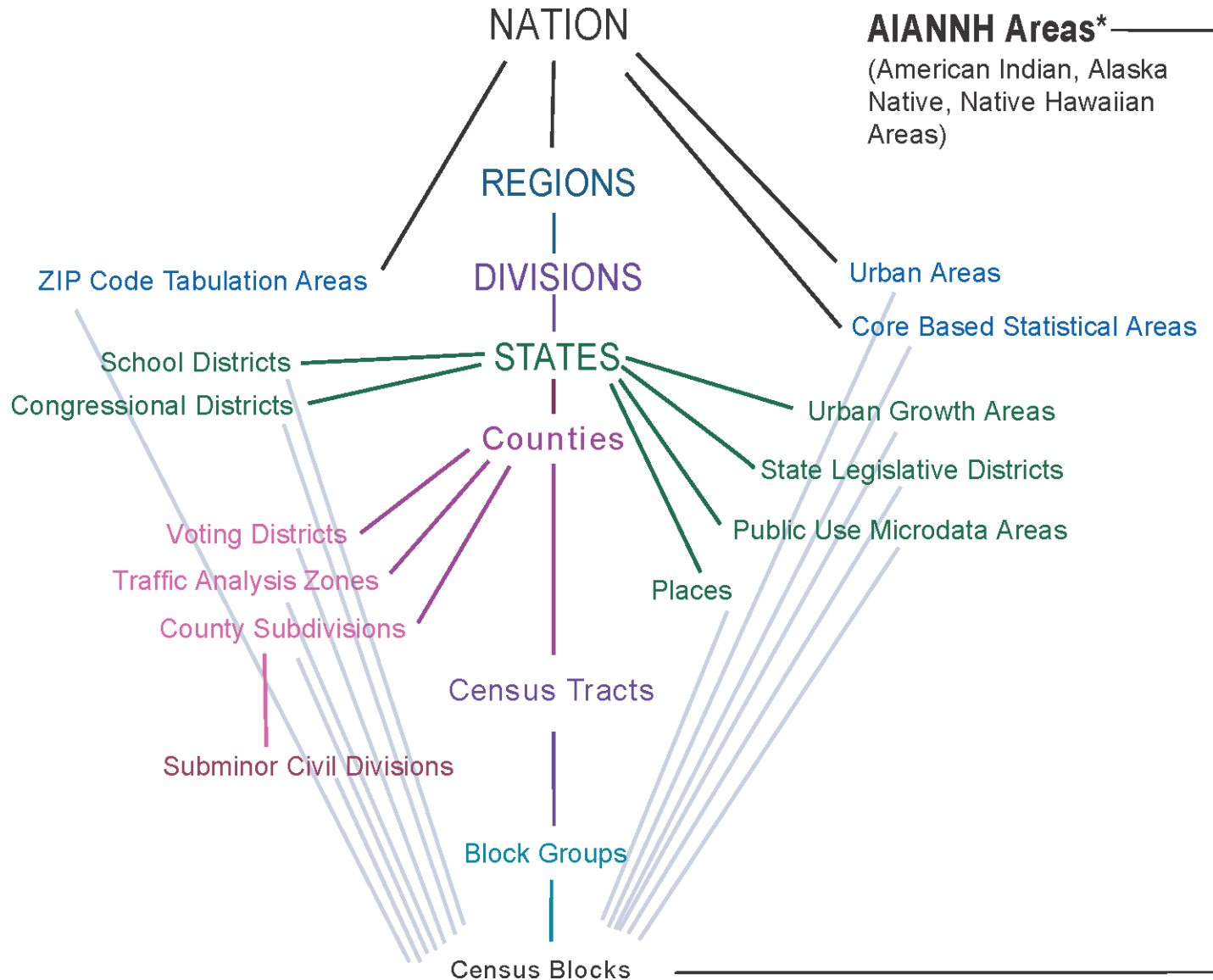
# SIPP Details II

- New construction frame provides coverage of structures for which building permits have been issued since the last census
  - Updated annually, and sampled with increasing percentage as the time since the last census increases
  - In the new construction frame, half the clusters have four expected housing units and half have eight expected units
- Statistical analysis is used to estimate the probabilities that households or individuals in the target population will be found in a particular frame (source of the design weight)
- The unit frame, which is based on the most recent Census, covers about 80% of the target population in the SIPP by these estimates

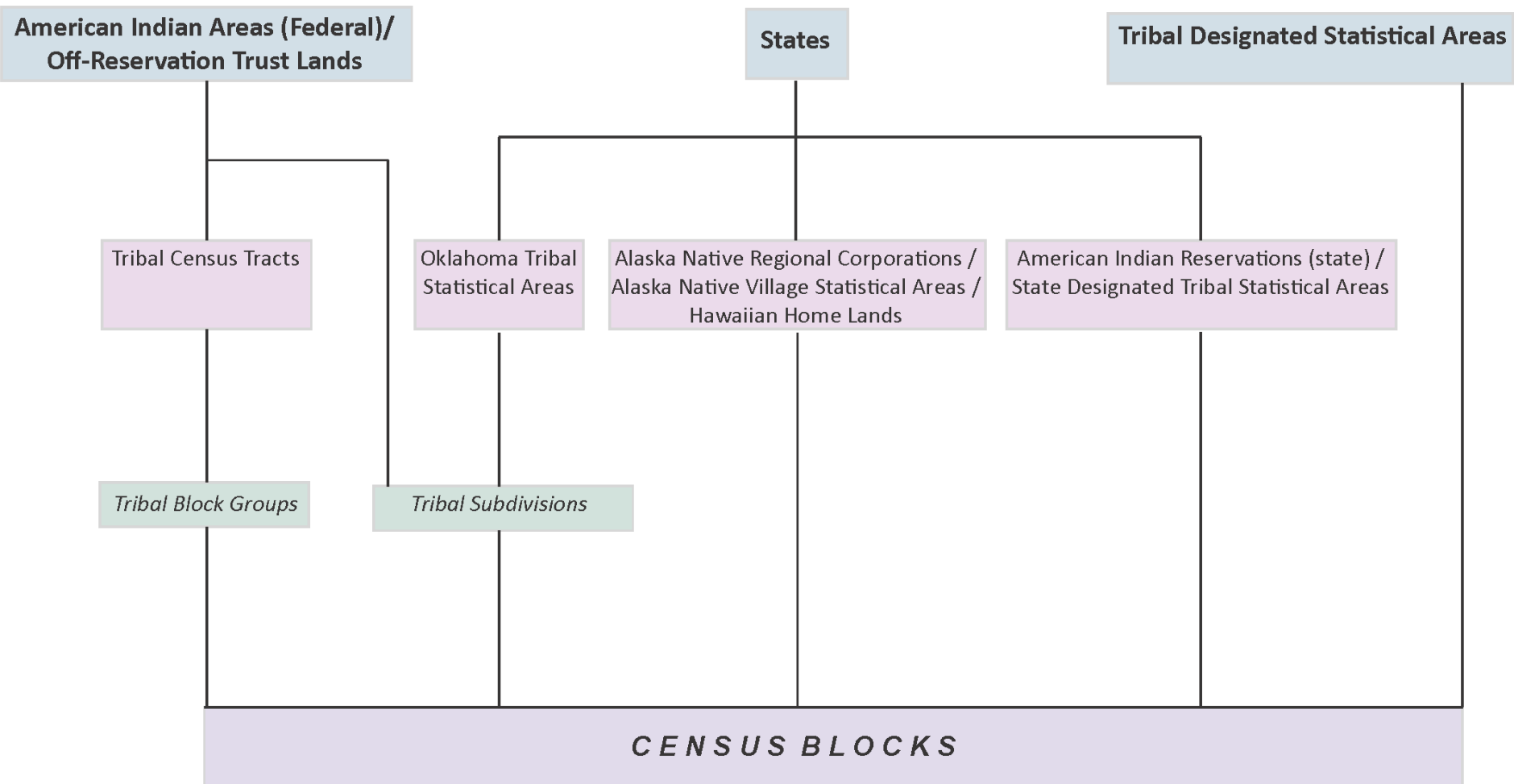
# Geography

- The geography of a sampling frame assigns to every latitude and longitude a fundamental geographic area
- Geographic entities can be assembled uniquely by aggregating geographic areas
- The basic geographic entity for the U.S. Census is the “block”

# Standard Hierarchy of Census Geography Entities



# Hierarchy of American Indian, Alaska Native, and Native Hawaiian Areas





# Entities

- In sampling frame development every geographic location (latitude and longitude) that contains a structure (natural or man-made) capable of originating economic activity is classified as a domicile, business, or both
- Entities are placed in the frame by declaring the target population to be humans beings (all domiciles including group quarters), economic (all businesses and service organizations), and government
- Notice that both for-profit and not-for-profit business activity are covered in the business entity scope

# Population Demographics

- Human populations are usually categorized by current living quarters when designing demographic sampling frames
- Distinguish between household living quarters and group living quarters
- Frames based on landline, mobile telephone or Internet service provider are not domicile base

# Household Living Quarter Definitions

- Housing unit: a house, an apartment, a mobile home or trailer, a group of rooms, or a single room occupied as separate living quarters, or if vacant, intended for occupancy as separate living quarters
  - Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall
  - For vacant units, the criteria of separateness and direct access are applied to the intended occupants whenever possible
- Household: A household includes all the people who occupy a housing unit as their usual place of residence

# Group Living Quarters

- Group quarters: all people not living in households
- There are two types of group quarters:
  - institutional (for example, correctional facilities, nursing homes, and mental hospitals)
  - non-institutional (for example, college dormitories, military barracks, group homes, missions, and shelters)

# Group Quarters Population Specifics

- Includes all people not living in households
- Includes those people residing in group quarters as of the date on which a particular survey was conducted
- Two general categories
  - the institutionalized population which includes people under formally authorized supervised care or custody in institutions at the time of enumeration (such as correctional institutions, nursing homes, and juvenile institutions)
  - the non-institutionalized population which includes all people who live in group quarters other than institutions (such as college dormitories, military quarters, and group homes). The non-institutionalized population includes all people who live in group quarters other than institutions.

# 10 minute break

# Business Entity Demographics

- Business entities have “establishments” as their basic unit
- Establishment: A business or industrial unit at a single location that distributes goods or performs services
- Establishments are collected into companies
- Business entity demographics separately track establishments (physical business locations) and companies (economic organizations owning establishments)

# Business Entity Demographics

- Company: (or “enterprise”) all the establishments that operate under the ownership or control of a single organization. A company may be a commercial business, service, or membership organization
- A company may consist of one or several establishments
- A company may operate at one or several locations
- A company may operate in one or more economic activities
- A company includes all subsidiary organizations, all establishments that are majority-owned by the company or any subsidiary, and all the establishments that can be directed or managed by the company or any subsidiary.



# Single-unit Companies (SU)

- Definition: Companies for which the location and the company are one and the same
- A single-unit business, service agency, or membership organization is one for which all the economic activity of the owner or owners is conducted at a single location
- Example: the “Shop Around The Corner” in the movie of the same name (or “You’ve got mail”)

# Multi-unit Companies (MU)

- Definition: companies that have more than one location
- A multi-unit business, service organization, or governmental agency is one for which the owners conduct economic activity at more than one physical location
- Example 1: all the manufacturing and management locations of the General Motors Corporation constitute a multi-unit company
- Example 2: all the service delivery locations of the Salvation Army constitute a multi-unit company

# Government Entities

- Definition: public entity created by the U.S. constitution, state constitutions or the statutes of a state
- In the United States these are divided into:
  - National government (U.S. constitution)
  - State government (state constitutions)
  - Local government (statutory entities created by states)
    - General purpose
    - Public school systems
    - Special districts

# Business and Government Entity Activity Classifications

- An entity that is engaged in economic or governmental activity may be classified in several ways
  - Ownership: the legal form of organization (public/private; corporate, partnership, sole proprietorship)
  - Activity: An industry is the most detailed category available in North American Industrial Classification System to describe business activities

# Business and Government Entity Activity Classifications

- NAICS provides hundreds of separate industry categories, unique categories that reflect different methods used to produce goods and services.
- Industry categories are used to classify, collect, process, publish, and analyze business statistics.
- [NAICS documentation](#)

# Demographic Sampling Frames

- Comprehensive mapping of location of the in-scope population with standardized geography
- Comprehensive list of domiciles in the geographic area covered
- Characteristics of inhabitants of the domiciles (stratifying variables)

# Coverage

- Domicile address lists are collected from multiple sources for households and group quarters
- The U.S. Census Bureau collects domicile addresses into the MAF (Master Address File)
- The U.S. Census Bureau collects physical locations into a set of files known as the TIGER (Topologically Integrated Geographic Encoding and Referencing system)
- [Census geography resources](#)

# Refreshing and Unduplication

- Refreshing a demographic sampling frame consists of collecting information on new housing units (households or group quarters) and purging information on housing units that have been taken out of service
- Unduplication is the process of expending resources to reduce the probability that an entity is present more than once in a sampling frame



# More about Unduplication

- When a demographic sampling frame is refreshed, addresses are added from multiple sources (tax records, new construction permits, *etc.*)
- The same address may appear more than once
- Some locations may appear to have no domiciles located on them but actually contain households or group quarters

# Economic Sampling Frames

- Comprehensive mapping of location of the in-scope activity with standardized geography
- Comprehensive list of addresses of business and government establishments in the geographic area covered
- Economic activity and size measures for the entity (stratifying variables)
- Updating of organizational structure (by survey)

# Coverage

- Entity address lists are assembled from previous economic censuses, tax records and surveys of company organization
- The U.S. Census Bureau collects business establishment information into the Employer Business Register and the Non-employer Business Register
- A separate register is maintained for government establishments
- The Bureau of Labor Statistics collects business establishment information into its Current Employment Statistics (CES) frame from information collected by state departments of employment security (ES-202)

# Frame Development

- Both business frames must collect information about the birth and death of new establishments
  - Census Bureau: [Report of Organization Survey](#) between Economic Censuses, responses on periodic surveys
  - BLS: [Quarter 1 of the QCEW and state Labor Market Information offices](#)
- This is complicated by the reliance on administrative reports (tax and unemployment insurance reports) and sample surveys

# Basic Relations Connecting Frames

- Geographic relations
- Business relations

# Geographic Relations

- Economic and demographic frames are directly connected through the use of common geographic identifiers
- A single location can be associated with household (demographic) activity, economic activity, both, or neither (undeveloped)
- No U.S. statistical agency maintains a single geographically integrated frame for households and businesses

# Business Relations

- Demographic and economic sampling frames can be connected by economic relations between the households and businesses
- Supplier-customer relations
- Employer-employee relations

# What is Sampling Frame Maintenance?

- The sampling frame or frame population defines the actual entities from the target population with a positive probability of inclusion in the survey or census
- Demographic and economic activity change the frame dynamically
- The biggest investment that statistical agencies and private firms make is the creation and maintenance of frame populations



# Housing Frame

- Swedish registers
- Census MAF/TIGER system (MTdb)
- Housing address list updates
- How do you combine the information?

# Register-based Population Censuses

- All personal addresses in Sweden are maintained in national registers
- The registers are updated by the individual when he/she moves using the national PIN
- All businesses and governmental activity use the registers to find the person
- The registers can be used as the sampling frame to conduct a census or survey
  - [Swedish register-based census 2005](#)
  - [German register-based census](#)

# Census Master Address File (Census 2010)

- Started with 2000 Master Address File
  - Based on 1990 Address Control File (ACG)
  - Add U.S. Postal Service Delivery Sequence File (DSF)
  - Un-duplication (one record per address)
- Updated with Census 2000 address improvements
  - 100% block canvas (independent of LUCA)
    - Physical visit of address; verification of use
    - Address confirmed or deleted
    - Missing addresses added if residential
  - Local Update of Census Addresses (LUCA)
  - Group Quarters Master File
- Spatial address database: Topologically Integrated Geographic Encoding and Referencing (TIGER) developed for 1990 Census

# Census MAF (2)

- Local Update of Census Addresses (LUCA; independent of 100% canvas)
  - Voluntary program by county and local governments
  - Adds and deletes by address
  - Special statutory exception to Title 13 for address improvement
  - States were also authorized in 2010 Census to cooperate
- [LUCA final assessment 2010 Census](#)

# Census MAF (3)

- New construction
  - To include new construction since 100% canvas and LUCA
  - Updated DSF
  - Lists of new housing construction from local governmental units
- Arbitration of units found by LUCA not in 100% canvas (field verification)
- Unduplication of new construction list
- Field staff visits to determine Census status

# MAF/TIGER System (MTdb)

- Contains all historical MAF data and metadata
- MAF Units:
  - Housing unit (HU)
  - Group quarters (GQ)
  - Transitory location (TL)
  - Nonresidential
- Ongoing update efforts
  - USPS Delivery Sequence Files (DSF)
  - ZIP Move Engineering File
  - Locatable Address Conversion System (LACS)
- [Overview from census.gov](#)

# Employer Business Frame Populations

- Census Employer Business Register
- BLS Establishment Register
- Establishment births and deaths
- The problem of false births and deaths
- The problem of multiple activity codes

# Census Employer Business Register

- Frame population consists of businesses that file income tax returns
- Employers are identified from the tax forms
- Multiunit and single unit businesses are determined in the Economic Census
- Updates to the MU/SU classification are determined by the annual Report of Organization Survey
- [Overview from census.gov](#)



# Business Register (2)

- Frame maintenance is conducted using weekly strips from the IRS master business returns
- Information is acquired with a lag that corresponds to the difference between filing deadlines and accounting periods
- Many different business tax forms are used

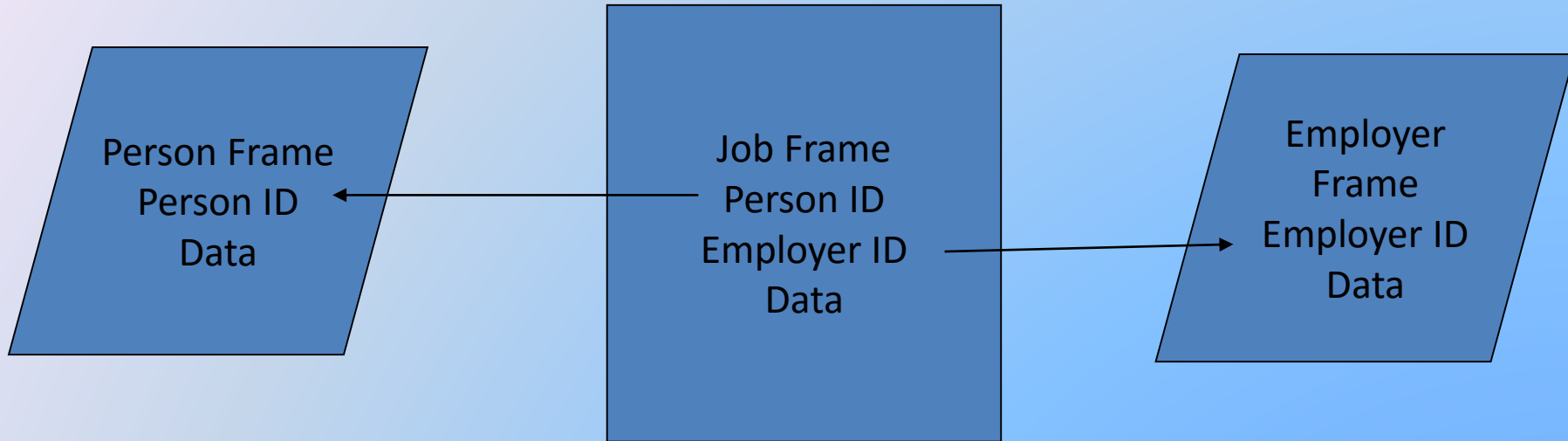
# Business Register (3)

- To create a frame population from the Business Register (employer businesses)
  - Define the target population by activity and date
  - Select the establishments from the SU and MU business registers that meet the target population definition
  - Eliminate inactive establishments
  - Edit size measures (employment, payroll, sales)

# Employment/Job Frame Populations

- A job is a relation between an employer and an employee
- Target population of jobs depends on definitions of employer and employee
- Frame population must be constructed from legal employment definitions
- No U.S. agency maintains a sampling frame for jobs
- Closest is the LEHD Infrastructure File System Employment History File at the Census Bureau

# Job Frame Populations



- Dynamic frames
- Defining an employer
- Defining an employee

# Job Frame Populations

- The problem of integration on the employer side
  - The job frame does not define a complete employer population frame, even if it is universal, unless the activity definitions for the employer and the employee report match
- The problem of integration on the employee side
  - The job frame does not define a complete individual population frame, even if it is universal, unless the individual activity definition only includes employment (no unemployment or non-participation)

# Geographic Link Frame

- Transportation network frames
- Defining a workplace location
- Defining a residence location
- Frame maintenance for the workplace-residence pair

# Transportation Network Frame Populations

- Potential target populations:
  - Trips to work on a particular date
  - Trips to purchase a good or service on a particular date
  - Leisure trips on a particular date
- Potential frame populations
  - Reported place of work on Decennial Census
  - Residential and employer address information in a job frame

# Defining a Workplace Location

- Latitude and longitude of the trip origin
  - With confidentiality protections, this origin definition is too fine to protect the identity of the traveler for a public use file
  - Frame population can be constructed from small geographic areas that include lat/long boundaries
- Census 2000/ACS use tracts and Traffic Analysis Zones (TAZ) [[guidelines](#)]
- OnTheMap uses blocks [[definitions](#)]



# Defining a Residence Location

- Latitude and longitude identify the location of the business exactly
  - Same concerns for public use files as in residential address
  - Frame populations constructed by assigning lat./long. to a particular unique area
- Census 2000, ACS: TAZ, tracts
- OnTheMap: tracts

# Transportation Frame Population Maintenance

- The complete frame population is defined by all possible origin/destination pairs in the geography chosen
- Census 2000 frame population constructed from the long form questions; not currently updated
- ACS frame population constructed from the 5-year composite tables
- OnTheMap frame population constructed from the job frame; updated annually for workplace-residence modeling

# Summary

- There are many important details associated with frame definitions and maintenance
- These details directly affect the quality of internal statistical agency data files that are directly used for research
- The connection between the operational procedures and the publication data will be examined in upcoming classes