

INFO 7470/ILRLE 7400  
The Decennial Census of  
Population and Housing

John M. Abowd and Lars Vilhuber  
February 15, 2011

# Navigating the (Virtual)RDC

## VirtualRDC

- Virtual thin client: NX on your desktop
- KDE desktop a choice
- Think of the Windows-C drive equivalent, RDC-root, “/ssgprojects/virtualrdc/”
- Fewer software choices, but SAS available

## Census RDC

- Physical thin client in the RDC, running NX
- KDE desktop (always)
- Think of the Windows-C drive equivalent, RDC-“root” :  
“/”
- More software choices, but SAS is prime



TACC-CAC-Training



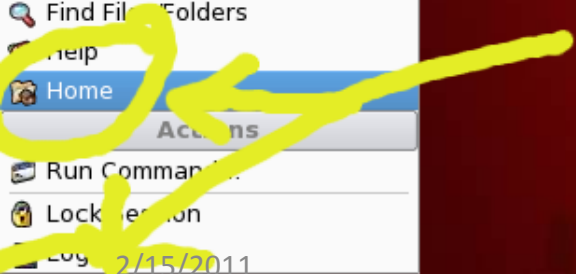
Trash

Most Used Applications

- Terminal
- Home
- SAS 9.2
- Kate
- Firefox Web Browser

All Applications

- Administration
- Development
- Graphics
- Internet
- Settings
- System
- Utilities
- Lost & Found
- Control Center
- Find Files & Folders
- Help
- Home
- Accessories
- Run Command
- Lock Session



lv39 - Konqueror

Location Edit View Go Bookmarks Tools Settings Window Help

Location: /home/lv39

- Home Folder
  - bin
  - bin.local
  - calpont-infinidb-1.0.2
  - Desktop
  - extras
  - R
  - sasuser.v92
  - synthesizer
  - tmp
  - Workspace
  - Workspace-non-encyr

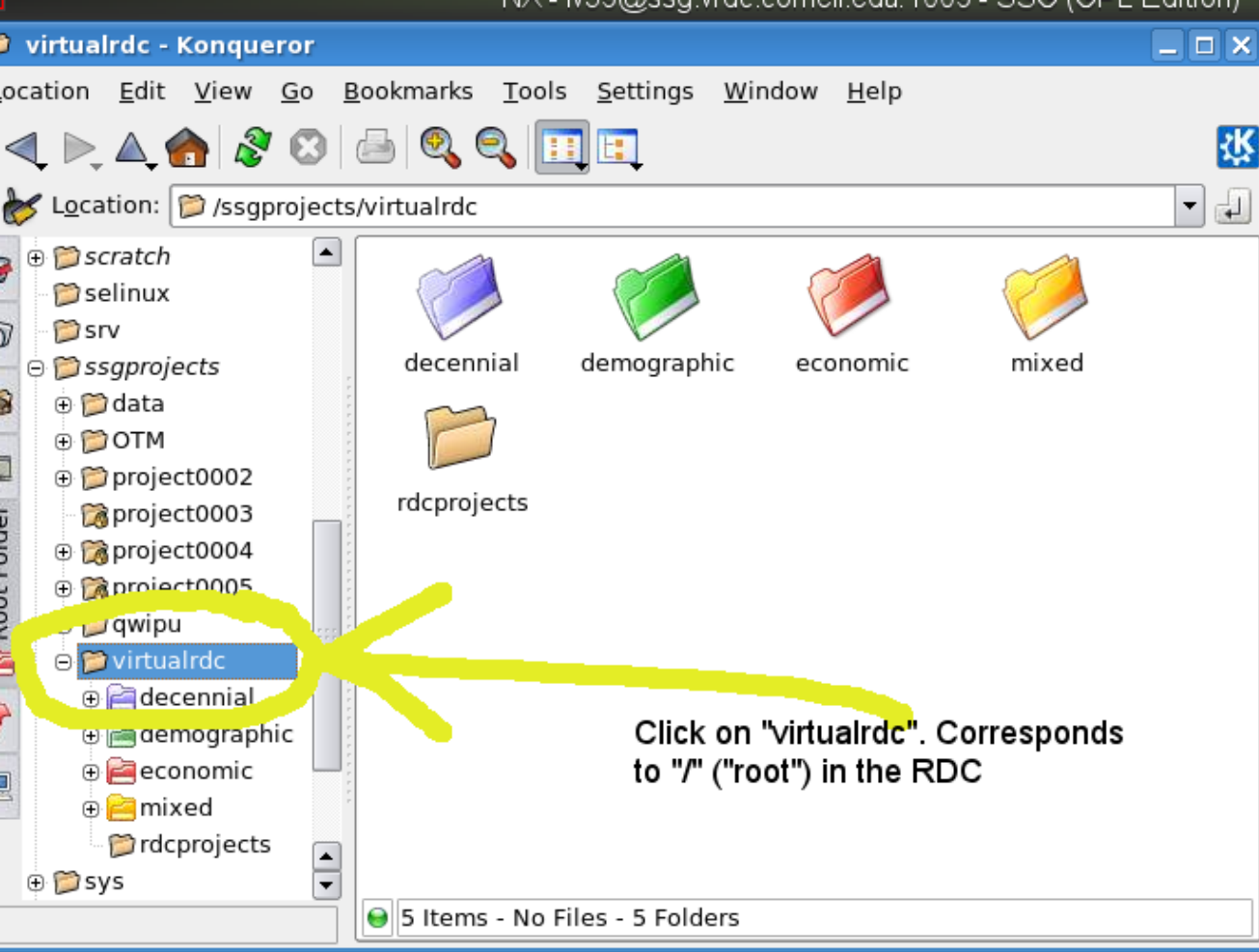
bin	bin.local	Desktop	extras
R	sasuser.v92	synthesizer	tmp
Workspace	Workspace-non-encrypt...	account_create	account_createSV
autoexec.sas	calpont-infinidb-1.0...	check_interactive_...	check_interactive_...

30 Items - 20 Files (15.8 MB Total) - 10 Folders

lv39 - Konqueror

Location: /home/lv39

30 Items - 20 Files (15.8 MB Total) - 10 Folders



Click on "virtualrdc". Corresponds to "/" ("root") in the RDC

# Outline

- Who's afraid of the VRDC, VRDC, VRDC ...?
- Some review
- Confidential Micro-data
- Structure of the data
- Using the short-form data
- Using the long-form data
- Comprehensive metadata

# Basic Definitions

- Household living quarters
- Group living quarters
- Census geography
- Out-of-scope population
- Frame population
- Geography
- Population demographics



# 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.

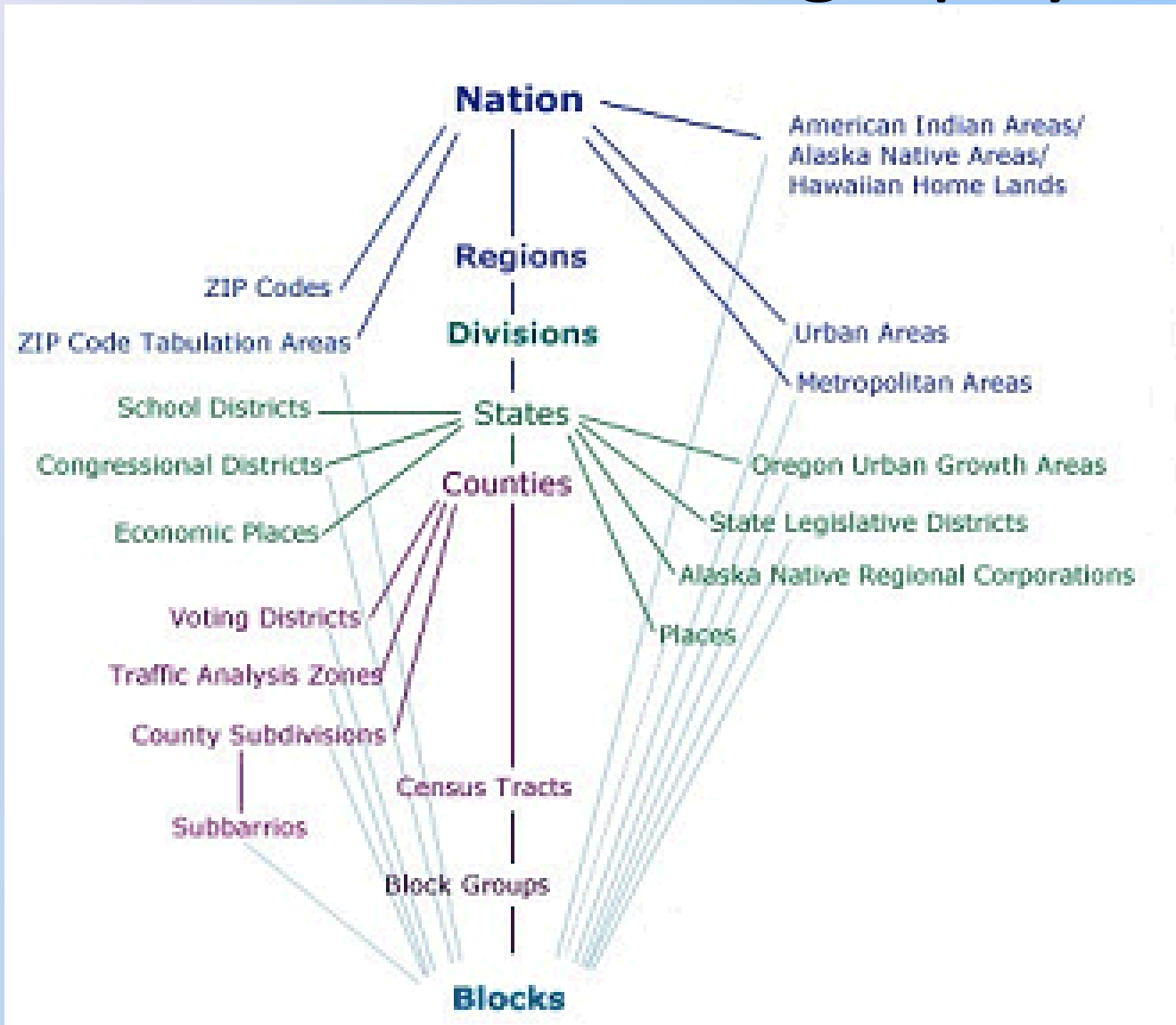
# Group Quarters Population

- 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.

# 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”

# U.S. Census Geography



# Confidential Micro-data

- Historical censuses
  - 1970-2000
  - Short form (called the “hundred percent”)
  - Long form (called the “sample” data)
- Naming conventions
  - T13[state][census year][dataset][record type]
  - State: FIPS/Postal code
  - Census year: 1970-2000
  - Dataset: h=100 percent; s=sample
  - Record type: g=block; h=household; p=person

# Using the Short Form Data Called “Hundred Percent Data”

- Documentation on the SSG
  - /ssgprojects/virtualrdc/decennial
  - See file: Zero Obs Files on VirtualRDC.doc
- Universe of persons
  - Use record type=p
- Universe of households
  - Use record type=h
- Universe of blocks
  - Use record type=g
- At this time, no group quarters data are available

# Hierarchical Structure

- Examples from the 1990 Census
- Others are similar
- On block-level record
  - Use the Geographic ID
- On the household record
  - Link the block-level data with Geographic ID
  - Link person-level data with the Household ID
- On the person record
  - Link the household-level data with Household ID
  - Link the block-level data with Geographic ID

# Using the Data

- Universe is Non-group-quarters population
- Person and household records are usually tabulated without weights
- In general, you can prepare a summary table (means, etc.) for data below the state-level to include in your papers (disclosure avoidance restriction)
- Most applications use detailed geocoding



# Geospatial Analysis

- We will cover this topic in a later lecture
- In general, you will need to prepare latitude and longitude for block-level centroids from a public source and request that the file be placed on your RDC project space
- There are tools on the VirtualRDC to help with this

# Metadata and Codebooks

- In general, the variables are coded using the same values as in the public-use micro-data samples
- Comprehensive meta-data are on the ICPSR web site
  - Example 1990  
<http://dx.doi.org/10.3886/ICPSR21983>

# Using the Long-form Data

- Essentially the same as using the PUMS files
- More geographic detail
- Fewer coarsening edits
- Complete long-form sample
- Again, there are restrictions on the kinds of summary tables that can be produced. None below the state level

INFO 7470/ILRLE 7400  
Economic Censuses and the  
Business Register

John M. Abowd and Lars Vilhuber  
Original Lecture by Jim Davis (Boston RDC)  
February 15, 2011

# Outline

- Economic Census (1997 example)
- Business Register (1997 example)
- Record Structure
- Identifiers
- Creating Establishment Analysis Files
- Creating Company (Alpha) Files
- Using Bridges to Other Data

# Economic Census (EC) 1997

- Target population:
  - Employer and non-employer establishments in covered industries (essentially everything except agriculture and government)
- Methods:
  - Large employers and a sample of small employers covered by mail questionnaire
  - Remaining establishments estimated from administrative records (mostly tax returns)

# Economic Census Files

Industry		63, 67	72	77	82	87	92	97	02, 07
CMI	Mining					X	X	X	X
CCN	Construction		X	X	X	X	X	X	X
CMF	Manufacturing	X	X	X	X	X	X	X	X
CUT	Transportation and Utilities					X	X	X	X
CW H	Wholesale			X	X	X	X	X	X
CRT	Retail			X	X	X	X	X	X
CFI	Finance, Insurance and Real Estate						X	X	X
CSR	Services			X	X	X	X	X	X
AUX	Auxiliaries			X	X	X	X	X	**

# Questionnaire Forms

- 1997 Economic Census Forms  
<http://www.census.gov/epcd/www/ec97form.html>
- 2002 Economic Census Forms  
<http://bhs.econ.census.gov/econhelp/resources/>
- History of the 1997 Economic Census  
<http://www.census.gov/prod/ec97/pol00-hec.pdf>
- Procedural History of the 2002 Economic Census  
<http://www.census.gov/prod/ec02/ec02-00r-hist.pdf>
- Accountants' Guide to the 2002 Economic Census  
<http://www.census.gov/epcd/ec02/pdf/acctsguide.pdf>
- University Library Government Documents



# Employer Business Register (BR) 1997

- Target Population:
  - Employer establishments in the same industries as are covered by the Economic Census
- Methods:
  - Continuously updated database of establishments divided into multi-unit and single-unit businesses

# Record Structure

- Economic Census
  - All establishments whether from mail questionnaire or administrative record
  - Separate files for Construction, Manufactures, Mining, Retail Trade, Services, (Transportation, Communications, and Utilities), Wholesale Trade
  - Example files are from Census of Manufactures

# Record Structure

- Employer Business Register Single-units (SU)
  - One record for each single unit establishment
  - One record, called a submaster, for each multiunit company
- Employer Business Register Multi-units (MU)
  - One record for each establishment for each multi-unit
  - Company Organization Survey

<http://bhs.econ.census.gov/2002forms/nc99002.pdf>

# Identifiers

- Entity Identifiers
  - Census File Number (CFN)
  - Employer Identification Number (EC: EI; BR: EIN)
  - Permanent Plant Number (PPN)
  - Census Alpha (EC: EIALPHA; BR: derived)

# Census File Number (CFN)

- Used for both the EC and BR as the sort order and main index for the file
- Always Character 10 (\$10.)
- For Single-units
  - First character “0”
  - Last 9 characters Federal Employer Identification Number (EIN)
- For Multi-units
  - First character nonzero
  - First six characters Census Alpha
  - Last four characters establishment ID

# Employer Identification Number

- Taxable (legal) entity identifier
- Always Character 9 (\$9.)
- For SUs, equivalent to CFN (prepended with '0') and unique
- For MUs, applies to the owning entity
  - An Alpha (see below) may be associated with multiple EINs

# Permanent Plant Number

- Longitudinal link based on CFN
- Always Character 10 (\$10.)
- Quality improves since inception in 1982
- Longitudinal links in Longitudinal Business Database are preferable

# Census Alpha

- Identifies the business that owns (50% or greater interest) the establishment for MUs
- EC and BR-SU
  - Character 6 (\$6.) and never has a leading 0
  - Called EIALPHA
- BR-MU
  - Character 10 (\$10.)
  - Called ALPHA
- Used to construct enterprise-level entities



# Business Register Redesign

- 2002 Business Register redesigned
- SURVU\_ID replaced CFN
- SURVU\_TYPE
  - MU/SU
  - ARU = Alternative reporting unit
  - SBM = EIN level reporting
- File Structure
  - Base/Misc
  - Line/Trailer – line code (LCODE) observations
- Historical identifiers (e.g. cfn, ppn) carried forward for continuers

# Identifiers

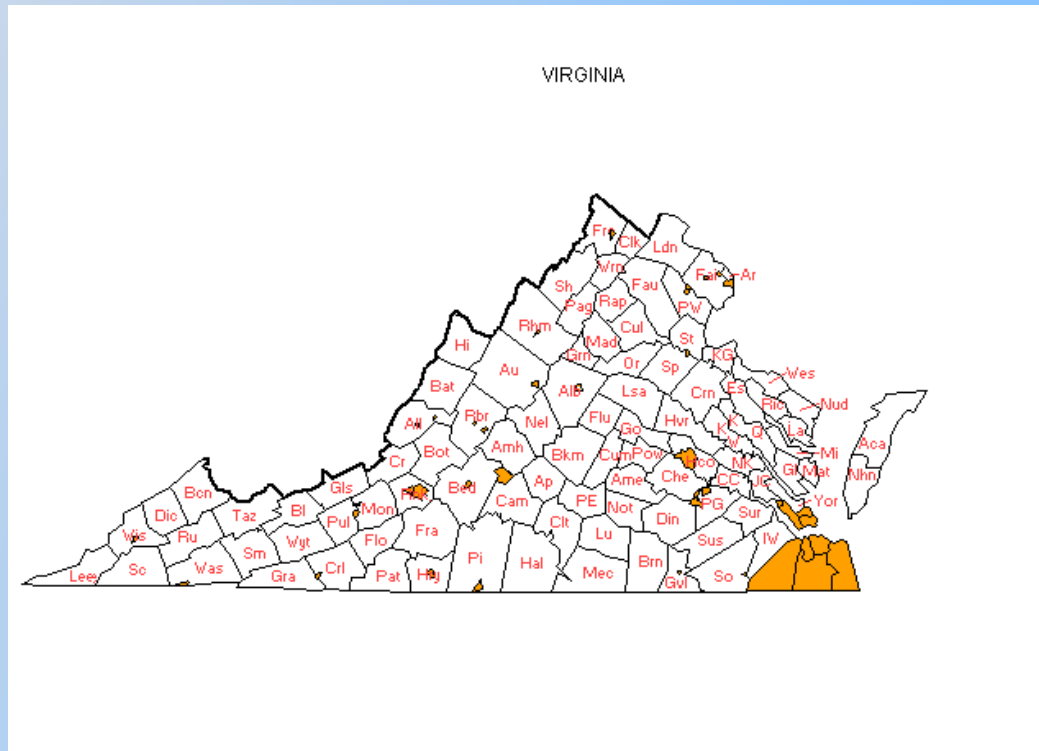
- Geography Identifiers
  - State identifiers
  - County identifiers
  - City identifiers
  - Full Census geography (BR only)
- Activity Identifiers
  - Industry Codes
  - Product Codes

# Geography Identifiers

- EC and BR
  - State (Census and FIPS)
  - County (Census and FIPS)
  - Consolidate Metropolitan Statistical Area (FIPS)
- BR only
  - Census Block
  - Zip
  - County Business Patterns Geography

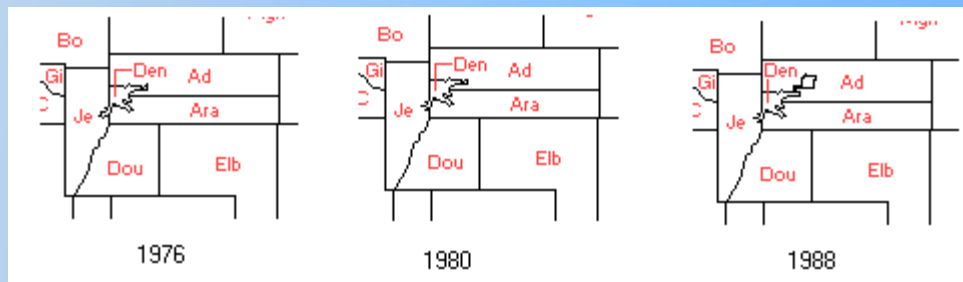
# Geography Identifiers

- Contemporaneous geographic definitions
- Virginia city-counties



# County Changes Since 1977

- La Paz, Arizona (created 1983) back into Yuma
  - replace fips=4027 if fips==4012
- Cibola, New Mexico (created 1981) back into Valencia
  - replace fips=35061 if fips==35006
- Washabaugh, South Dakota (merged 1979) back into Jackson
  - replace fips=46071 if fips==46131
- St. Genevieve, Missouri renumbered in 1982
  - replace fips=29186 if fips==29193
- Muskogee, Georgia renumbered in 1982
  - replace fips=13215 if fips==13510
- Denver



# Activity Identifiers

- NAICS
  - Full U.S. Industry code (Char 6)
  - Derived industry codes
- SIC
  - Full 1987 SIC (6-digits to product class code)
- Contemporaneous Industry Definitions
- NAICS to SIC bridge codes

# RDC 2002 NAICS Sectors

NAICS Sectors	RDC 1997/2002 Files
22, 48, 49	CUT (Utilities)
42	CWH (Wholesale)
44, 45, 72	CRT (Retail)
51, 54, 56, 61, 62, 71, 81	CSR (Services)
52, 53	CFI (FIRE)
55	CSR in 2002

Source: K. McCue 2006 Technical Note "Revised Business Census Files: 1977-2002 Census of Retail, Wholesale, Services, Communications/Utilities/Transportation, and FIRE"

# Finding Active Entities

- EC
  - Use sample weight ( $WT > 0$ )
- BR-SU
  - Exclude submasters ( $PDIV = 'M'$ )
  - Payroll or employment positive (see example)
- BR-MU
  - Exclude ghosts ( $ACT = 'G'$ )
  - Payroll or employment positive (see example)



# Creating Custom Entities

- Establishment-level files
  - EC: natural organization
  - BR-SU: exclude submaster records
  - BR-MU: exclude ghosts
- Company-level files
  - EC: use EIALPHA to find related establishments
  - BR: create compatible ALPHA10 in SU to get information on submasters to link to establishments on MU
- Pseudo-establishments
  - Combine establishments based on EIN, geography and activity