TERRA POPULUS

Integrated Data on Population and Environment

MPC
Minnesota Population Center

University of Minnesota
LIBRARIES

INSTITUTE ON THE
ENVIRONMENT
UNIVERSITY OF MINNESOTA
Driven to Discover

CIESIN
Columbia University

ICPSR
A PARTNER IN SOCIAL SCIENCE RESEARCH
INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

NSF

TerraPop Goals

Lower barriers to conducting interdisciplinary human-environment interactions research by making data with different formats from different scientific domains easily interoperable

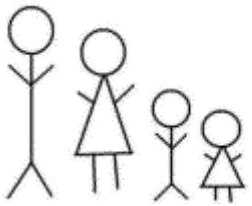
Provide an organizational and technical framework to preserve, integrate, disseminate, and analyze global-scale spatiotemporal data describing population and the environment.

Source Data

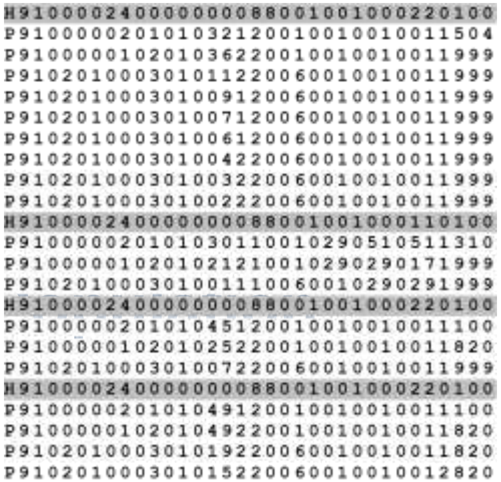
- DOMAINS & FORMATS
- POPULATION MICRODATA
- AREA-LEVEL DATA

Terra Populus Data Domains

Microdata

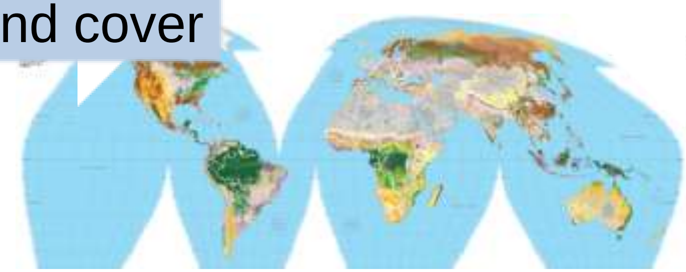


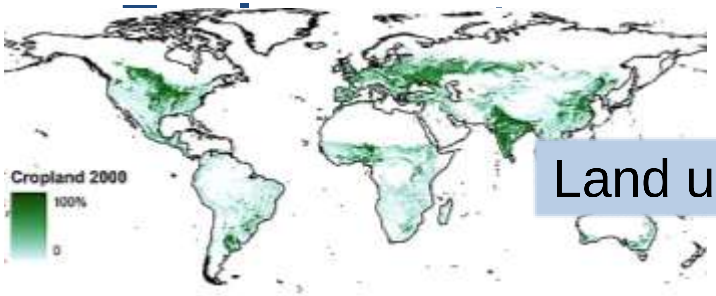
Individuals and households



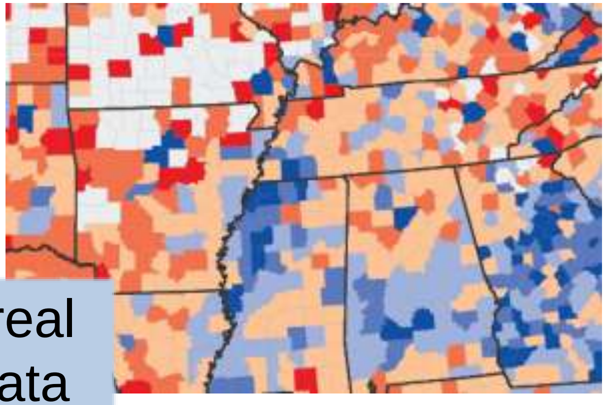
```
H910000240000000088001001000220100
P910000020101032120010010010011504
P910000010201036220010010010011999
P910201000301011220060010010011999
P910201000301009120060010010011999
P910201000301007120060010010011999
P910201000301006120060010010011999
P910201000301004220060010010011999
P910201000301003220060010010011999
P910201000301002220060010010011999
H910000240000000088001001000110100
P910000020101030110010290510511310
P910000010201021210010290290171999
P910201000301001110060010290291999
H910000240000000088001001000220100
P910000020101045120010010010011100
P910000010201025220010010010011820
P910201000301007220060010010011999
H910000240000000088001001000220100
P910000020101049120010010010011100
P910000010201049220010010010011820
P910201000301019220060010010011820
P910201000301015220060010010012820
```

Land cover

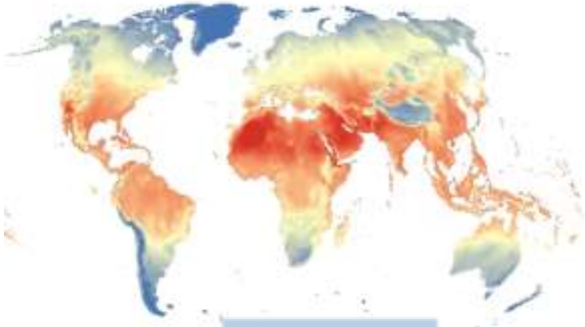




Land use

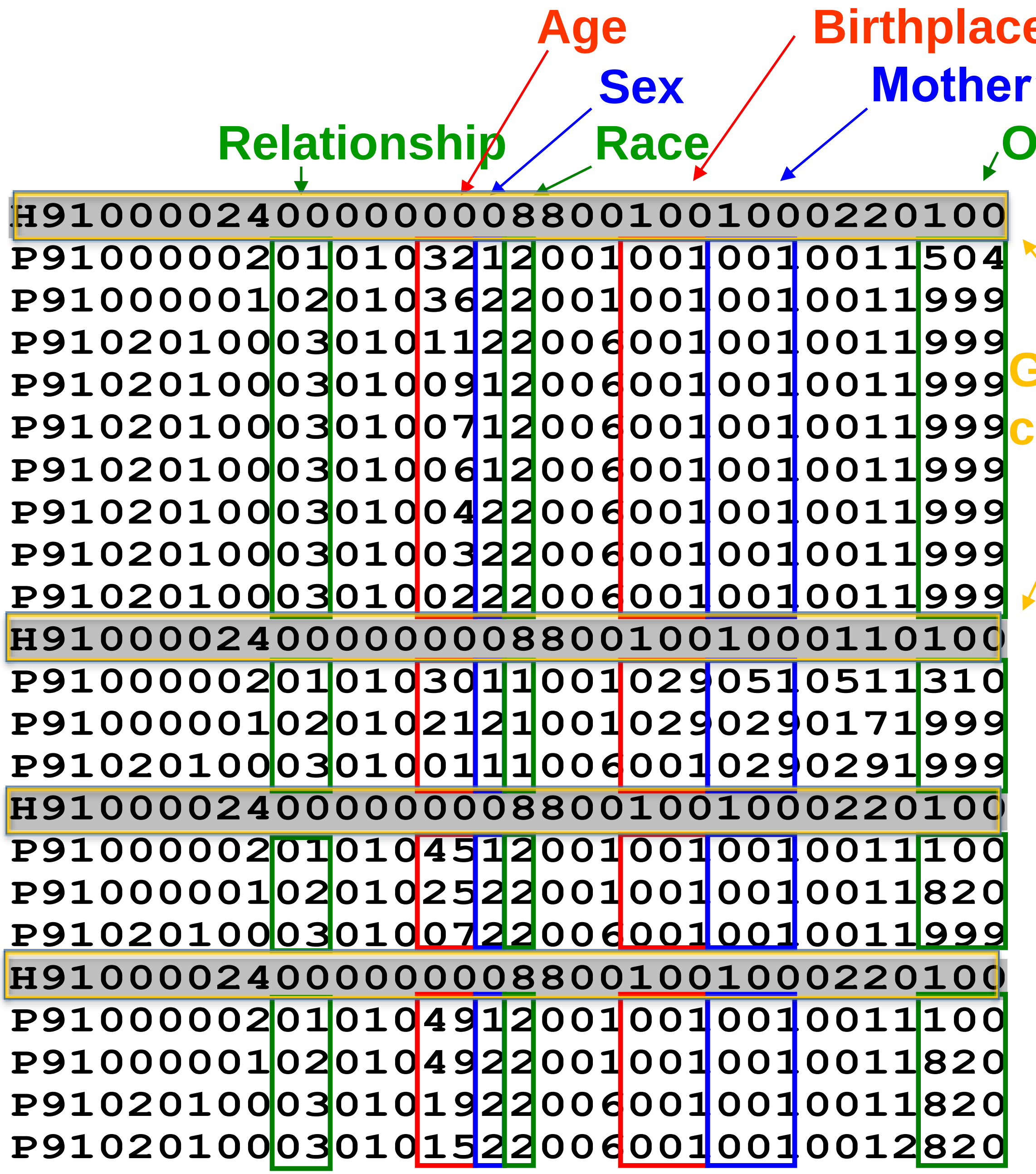


Areal Data



Climate

Population Microdata Structure

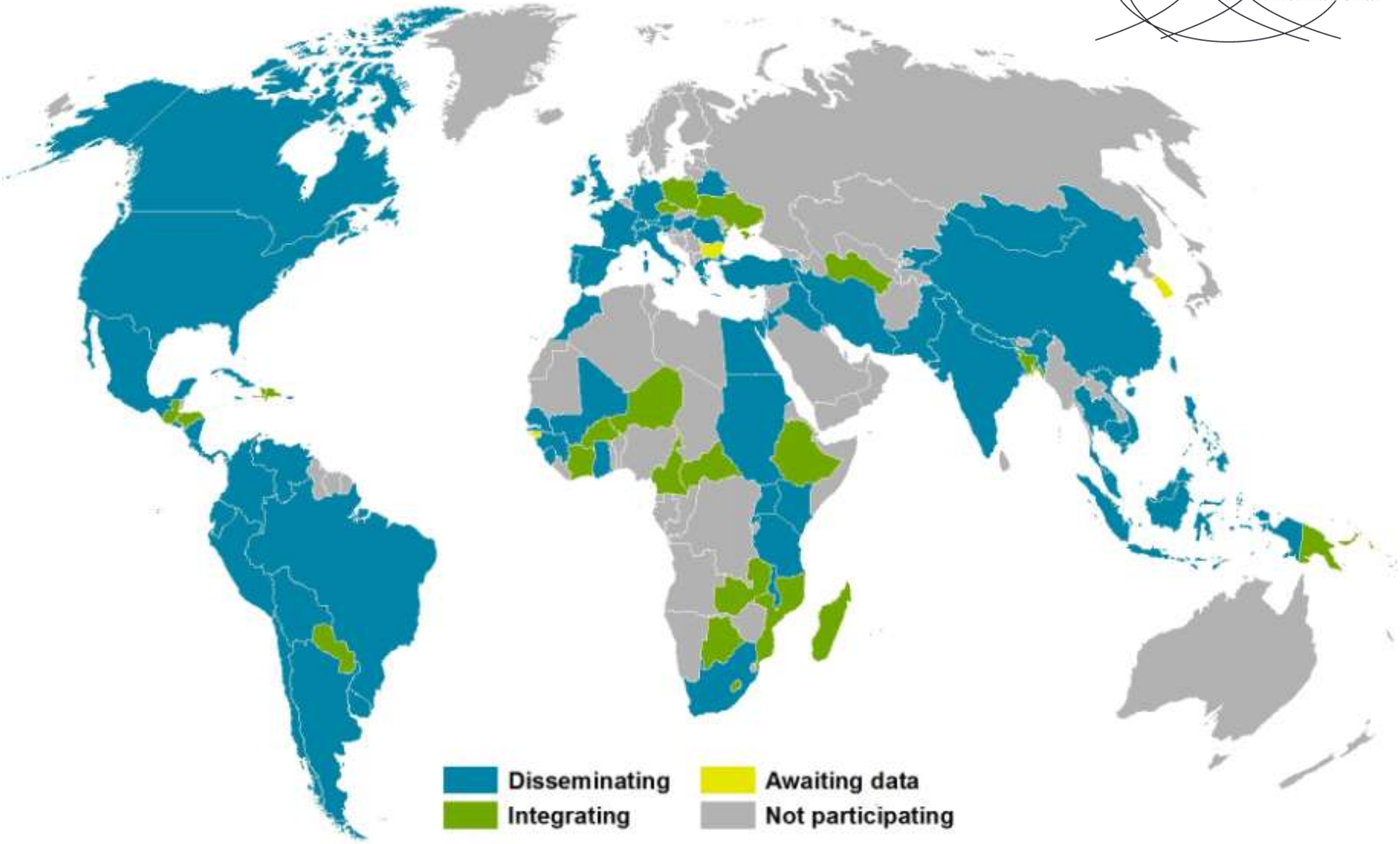


```
H910000240000000088001001000220100
P910000020101032120010010010011504
P910000010201036220010010010011999
P910201000301011220060010010011999
P910201000301009120060010010011999
P910201000301007120060010010011999
P910201000301006120060010010011999
P910201000301004220060010010011999
P910201000301003220060010010011999
P910201000301002220060010010011999
H910000240000000088001001000110100
P910000020101030110010290510511310
P910000010201021210010290290171999
P910201000301001110060010290291999
H910000240000000088001001000220100
P910000020101045120010010010011100
P910000010201025220010010010011820
P910201000301007220060010010011999
H910000240000000088001001000220100
P910000020101049120010010010011100
P910000010201049220010010010011820
P910201000301019220060010010011820
P910201000301015220060010010012820
```

Geographic and housing characteristics

- Rows
  - Household records
  - Person records within households
- Columns
  - Variables

Microdata Availability



Area-level Data Sources

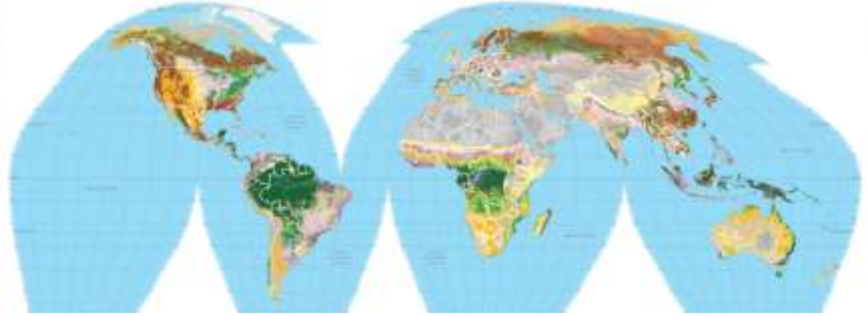
- Census tables, especially where microdata is unavailable
- Other types of surveys, data
  - Agricultural surveys
  - Economic surveys, data
  - Election data
- Legal information

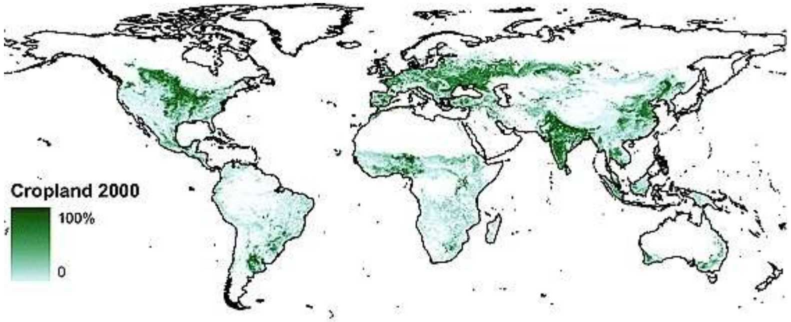
Environmental Data (Rasters)

TerraPop Prototype

- Land cover data from satellite images

  (*Global Land Cover 2000*)
- Agricultural land use data from satellites and government records (*Global Landscapes Initiative*)
- Climate data from weather stations (*WorldClim*)

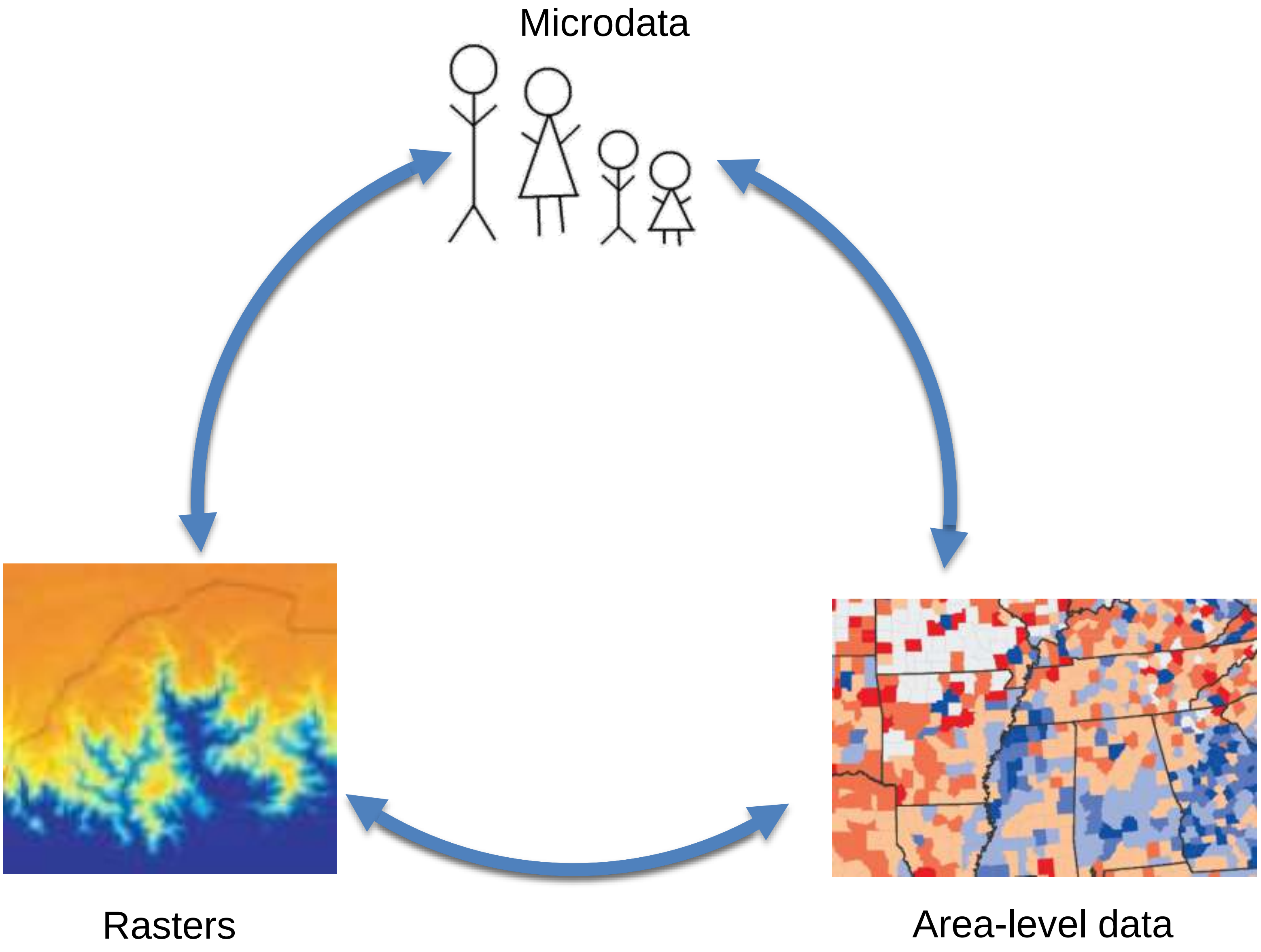




Location-Based Integration

MICRODATA ⇔ AREA-LEVEL ⇔ RASTER

Location-Based Integration



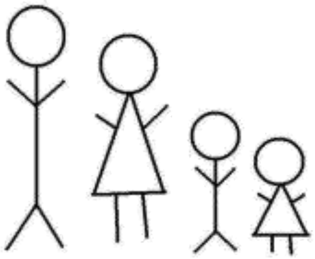
Microdata

Rasters

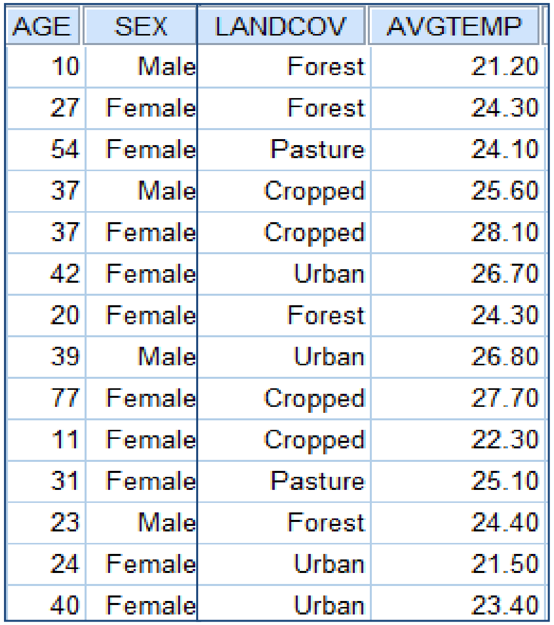
Area-level data

Location-Based Integration

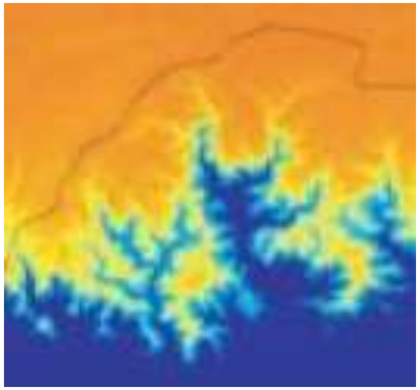
Microdata



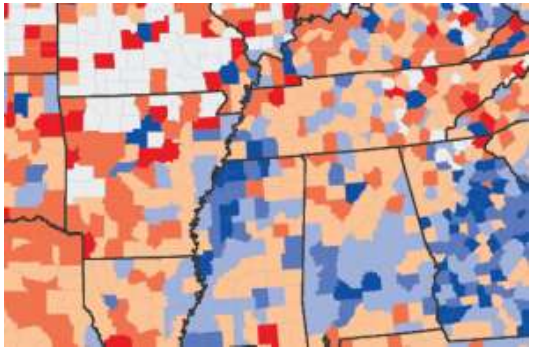
Individuals and households with their environmental and social context



AGE	SEX	LANDCOV	AVGTEMP
10	Male	Forest	21.20
27	Female	Forest	24.30
54	Female	Pasture	24.10
37	Male	Cropped	25.60
37	Female	Cropped	28.10
42	Female	Urban	26.70
20	Female	Forest	24.30
39	Male	Urban	26.80
77	Female	Cropped	27.70
11	Female	Cropped	22.30
31	Female	Pasture	25.10
23	Male	Forest	24.40
24	Female	Urban	21.50
40	Female	Urban	23.40



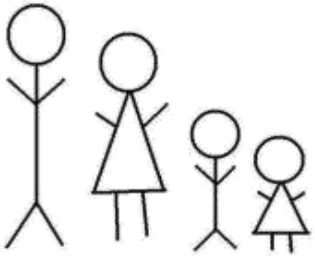
Rasters



Area-level data

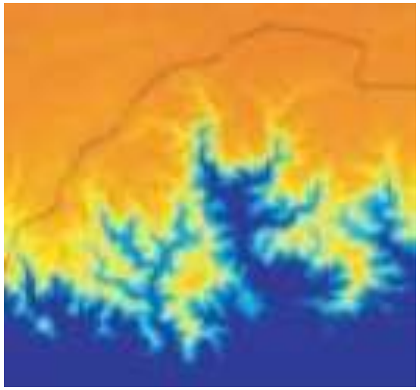
Location-Based Integration

Microdata

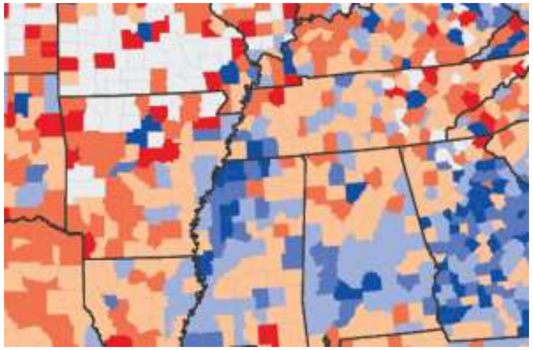


County ID	Mean Ann. Temp.	Max. Ann. Precip.	Rent, Rural	Rent, Urban	Own, Rural	Own, Urban
G17003100001	21.2	768	3129	1063	637	365
G17003100002	23.4	589	2949	1075	1469	717
G17003100003	24.3	867	3418	1589	1108	617
G17003100004	21.5	943	1882	425	202	142
G17003100005	24.1	867	2416	572	426	197
G17003100006	24.4	697	2560	934	950	563
G17003100007	25.6	701	2126	653	321	215

Summarized environmental and population characteristics for administrative districts



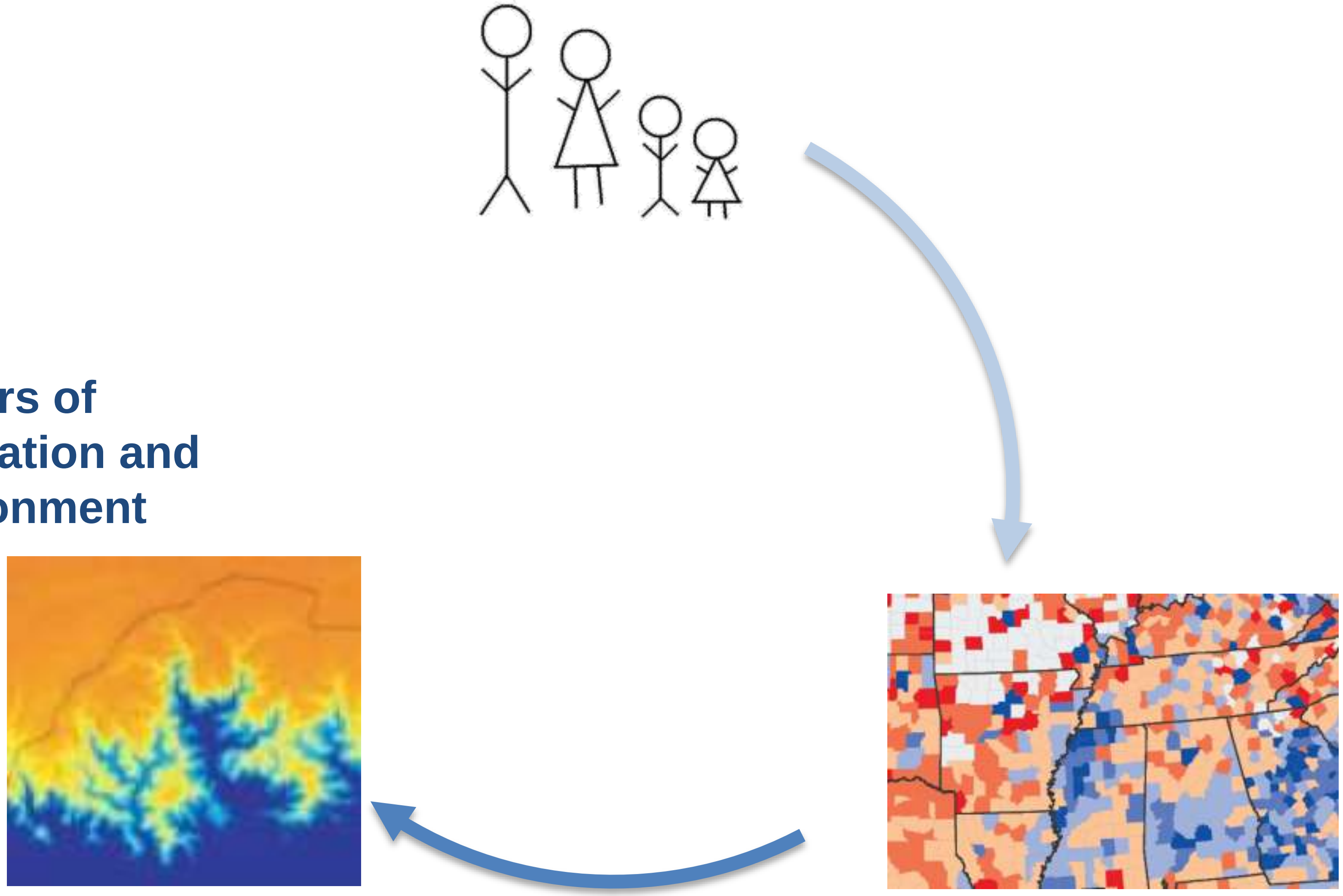
Rasters



Area-level data

Location-Based Integration

Microdata



Rasters of population and environment data

Rasters

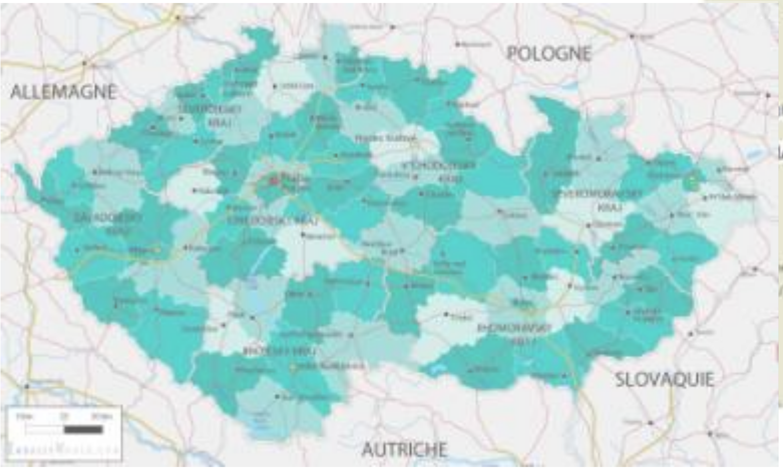
Area-level data

Boundaries are Key

- Linkages across data formats rely on administrative unit boundaries
- Particular needs
  - Lower level boundaries
  - Historical boundaries

NORWAY
NORWEGIAN SEA
NORRBOTTEN
VÄSTERBOTTEN
JÄMTLAND
VÄSTERNORRLAND
GÄVLEBORG
DALARNA
NORWAY
FINLAND
VÄRMLAND
UPPSALA
ÖREBRO
STOCKHOLM
SÖDERMANLAND
VÄSTRA GÖTALAND
ÖSTERGÖTLAND
JÖNKÖPING
KALMAR
GOTLAND
DENMARK
HALLAND
KRONOBERG
SKÅNE
BLEKINGE
BALTIC SEA
LITHUANIA

SWEDEN
BALTIC SEA
KALININGRAD RUSSIAN FED
LITHUANIA
POMORSKIE
WARMINSKO-MAZURSKIE
PODLASKIE
ZACHODNIO-POMORSKIE
KUJAWSKO-POMORSKIE
MAZOWIECKIE
WARSZAWA
WIELKOPOLSKIE
LUBUSKIE
ŁÓDZKIE
LUBELSKIE
ŚWIĘTOKRZYSKIE
DOLNOŚLĄSKIE
OPOLSKIE
ŚLĄSKIE
PODKARPACKIE
MAŁOPOLSKIE
CZECH REP
SLOVAKIA
UKRAINE



Administrative Unit Boundary Processing

- OBTAINING
- LINKING TO MICRODATA
- TEMPORAL HARMONIZATION
- REGIONALIZATION

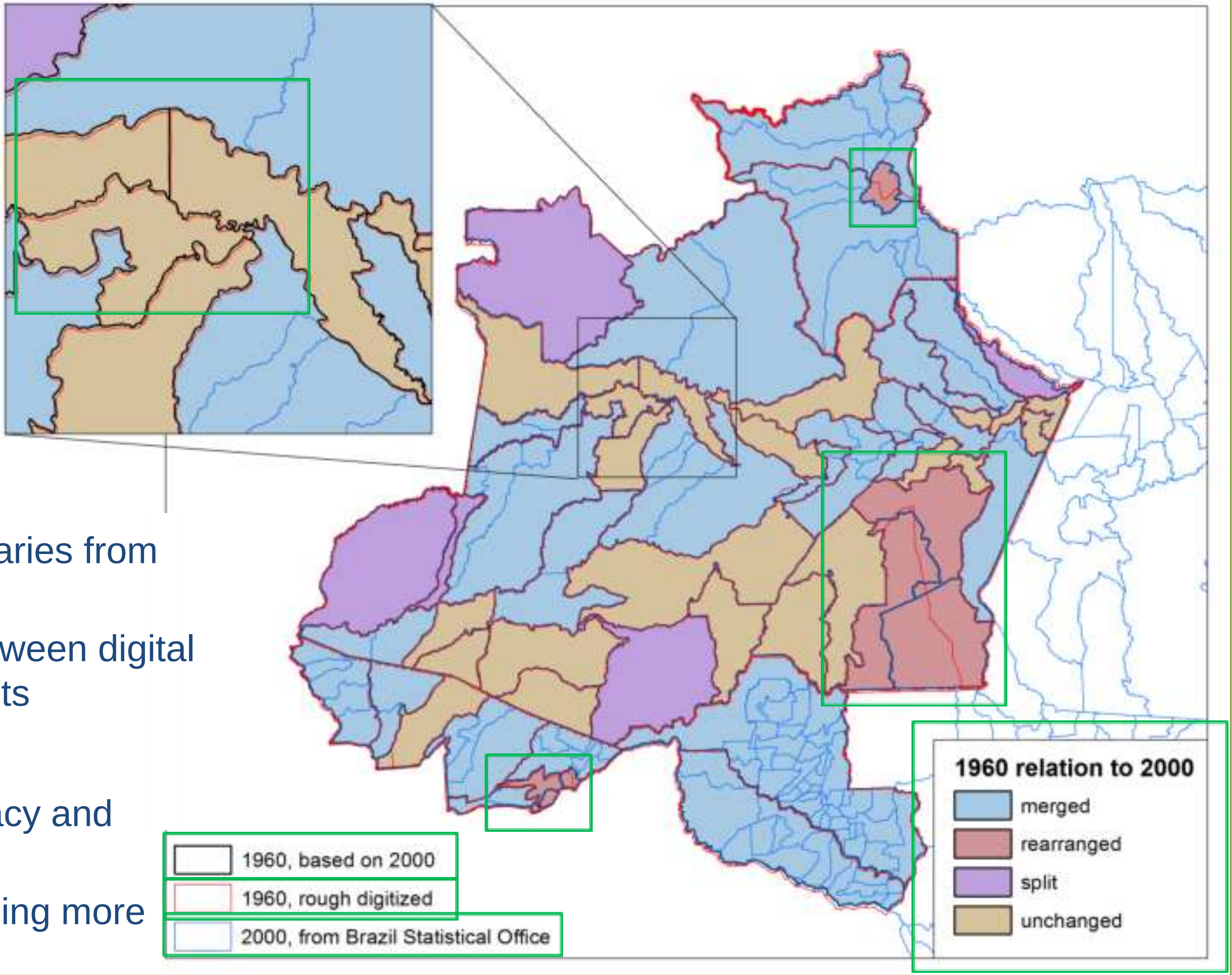
Obtaining Boundary Data

- Potential sources of digital data
  - National Statistical Offices
  - Global Administrative Areas data (e.g. SALB, GAUL)
  - Digitizing from images or paper maps
- Challenges
  - Lower level and historical data
  - Date mismatches with census data
  - Code matching to microdata

Digitizing Boundaries

Leveraging available digital data

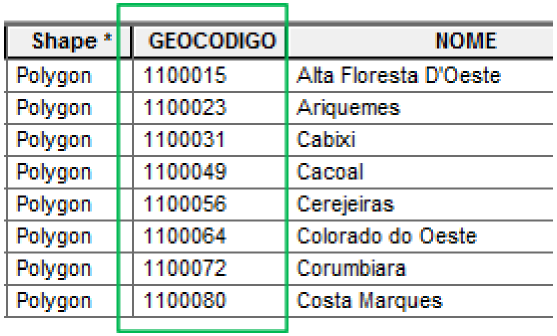
- Script input
  - Existing digital data
  - Rough digitized boundaries
- Script output
  - Relevant boundaries from digital data
  - Relationship between digital and digitized units
- Advantages
  - Preserve accuracy and detail
  - Flag areas needing more work



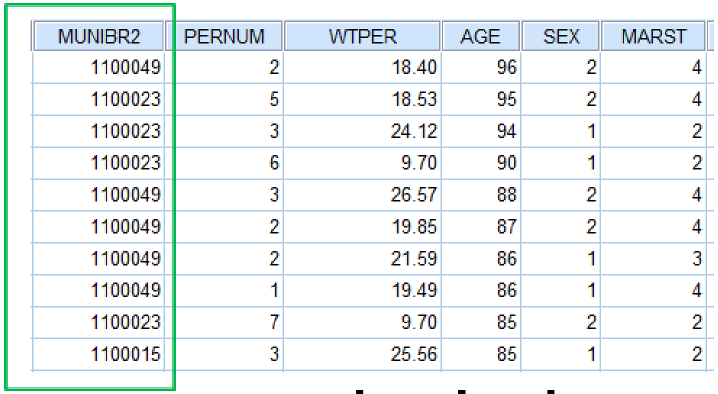
Code Matching

- Codes link boundaries to microdata records, connect people to places

Boundary shape attributes



Shape *	GEOCODIGO	NOME
Polygon	1100015	Alta Floresta D'Oeste
Polygon	1100023	Ariquemes
Polygon	1100031	Cabixi
Polygon	1100049	Cacoal
Polygon	1100056	Cerejeiras
Polygon	1100064	Colorado do Oeste
Polygon	1100072	Corumbiara
Polygon	1100080	Costa Marques



MUNIBR2	PERNUM	WTPER	AGE	SEX	MARST
1100049	2	18.40	96	2	4
1100023	5	18.53	95	2	4
1100023	3	24.12	94	1	2
1100023	6	9.70	90	1	2
1100049	3	26.57	88	2	4
1100049	2	19.85	87	2	4
1100049	2	21.59	86	1	3
1100049	1	19.49	86	1	4
1100023	7	9.70	85	2	2
1100015	3	25.56	85	1	2

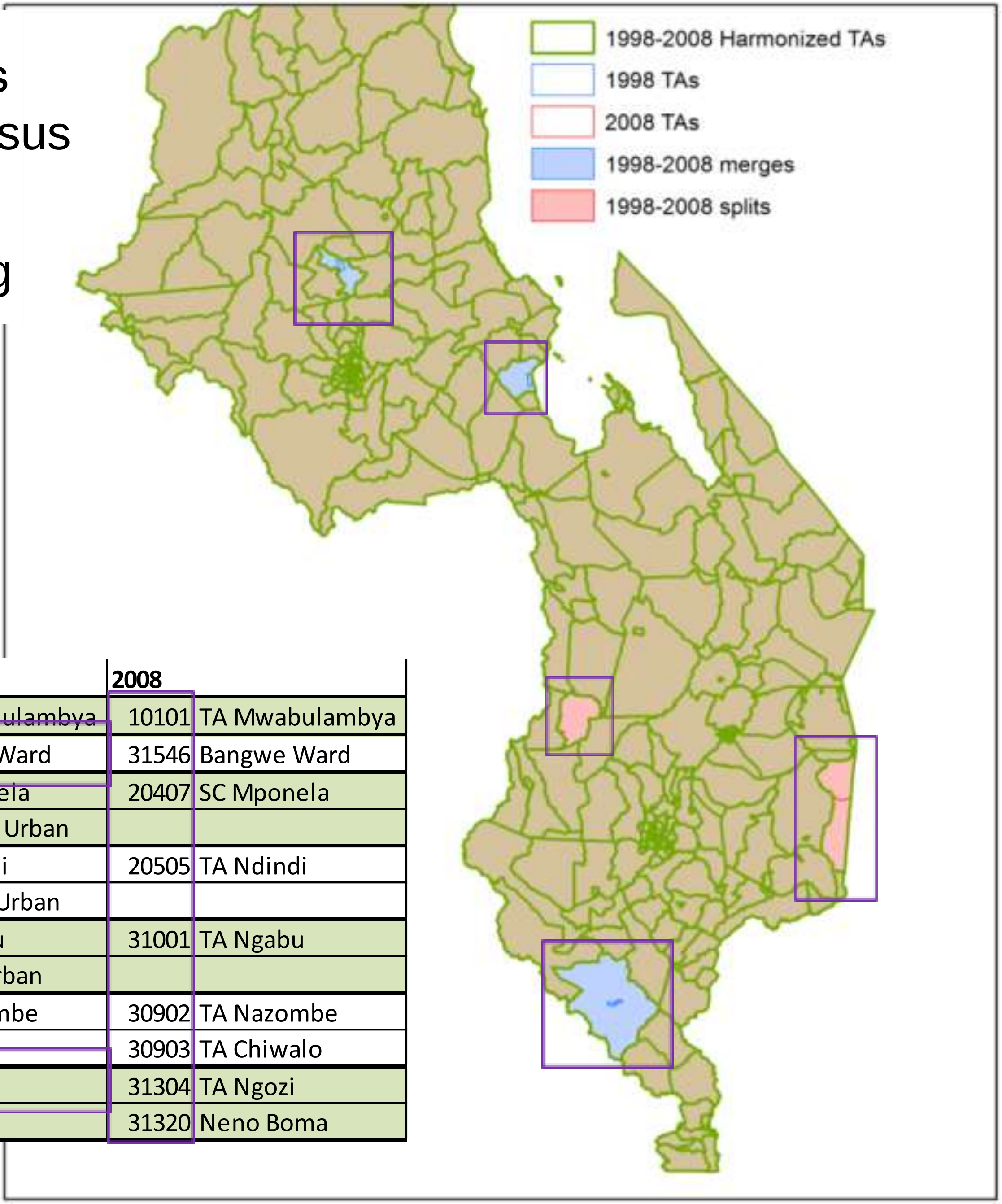
IPUMS microdata

- Boundary data may or may not include codes
- Approach
  - Name matching, when possible
  - Map observations – digitizing script captures codes
  - Research on boundary changes

Temporal Harmonization

- Purpose
  - Create consistent units for time-series analysis
- Top-down strategy
  - Start with first administrative level units
  - Harmonize 2nd level units within 1st level “containers”
- Script to create “least common denominator” units
  - Applicable when maps from multiple years are available
  - Creates aggregate units encompassing areas with boundary changes
  - Constructs source-harmonized crosswalk

- “Erase” interior boundaries applicable to only one census
- Apply harmonized codes
- Also aids in code matching



Crosswalk

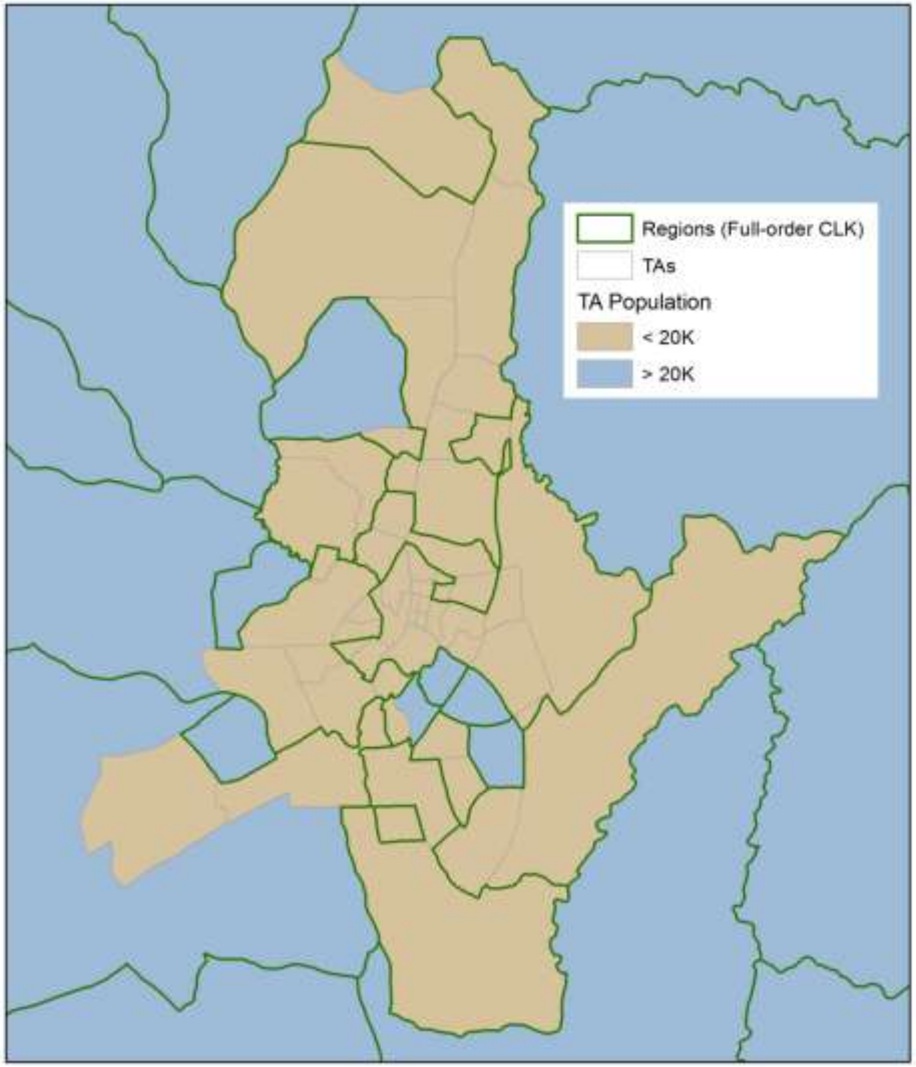
Harmonized		1998		2008	
10101	TA Mwabulambya	10101	TA Mwabulambya	10101	TA Mwabulambya
31546	Bangwe Ward	30546	Bangwe Ward	31546	Bangwe Ward
20407	Mponela	20407	SC Mponela	20407	SC Mponela
		20421	Mponela Urban		
20505	Ndindi and Chipoka Urban	20505	TA Ndindi	20505	TA Ndindi
		20521	Chipoka Urban		
31001	Ngabu	31001	TA Ngabu	31001	TA Ngabu
		31021	Ngabu Urban		
30902	Nazombe and Chiwalo	30902	TA Nazombe	30902	TA Nazombe
				30903	TA Chiwalo
31304	Ngozi and Neno Boma	30606	TA Ngozi	31304	TA Ngozi
				31320	Neno Boma

Regionalization

- Confidentiality concerns require minimum 20,000 population in each unit disseminated
- REDCAP tool
  - Constructs regions by combining units
  - Regions meet minimum population threshold
  - Contiguity constrained
  - Combines units that are similar in terms of a selected variable
- Currently in testing phase
  - REDCAP Algorithms and parameters
  - Optimization variables (e.g., pop. density, education, occupation)
  - Testing on Malawi TAs, Brazil 2000 municipios

Regionalization - Lilongwe, Malawi

- Units < 20K combined with neighbors to meet threshold
- Specific aggregation depends on
  - Optimization variable
  - Algorithm



Beyond Administrative Boundaries

- ARBITRARY BOUNDARIES
- RASTERIZATION

Arbitrary Boundaries

- Watersheds, buffers around features, etc.
- Near-term
  - Summarize rasters to user-supplied boundaries
  - Identify administrative units intersecting user-supplied boundaries

- Future
  - Reallocation based on uniform distribution assumption
  - Reallocation based on other assumptions

Rasterization

- Prototype - All cells in unit get the same value
  - Use lowest level units available
  - Rates only, not counts
- Future – Distribute based on ancillary data
  - Requires research on available methods
  - May provide as service – users select:
    - Ancillary data
    - Weights
    - Spatial distribution parameters