Data Challenges for the Development Agenda: Observations from IPUMS

United Nations Expert Group Meeting on
Strengthening the Demographic Evidence Base for The Post-2015 Development Agenda

New York
5-6 October 2015

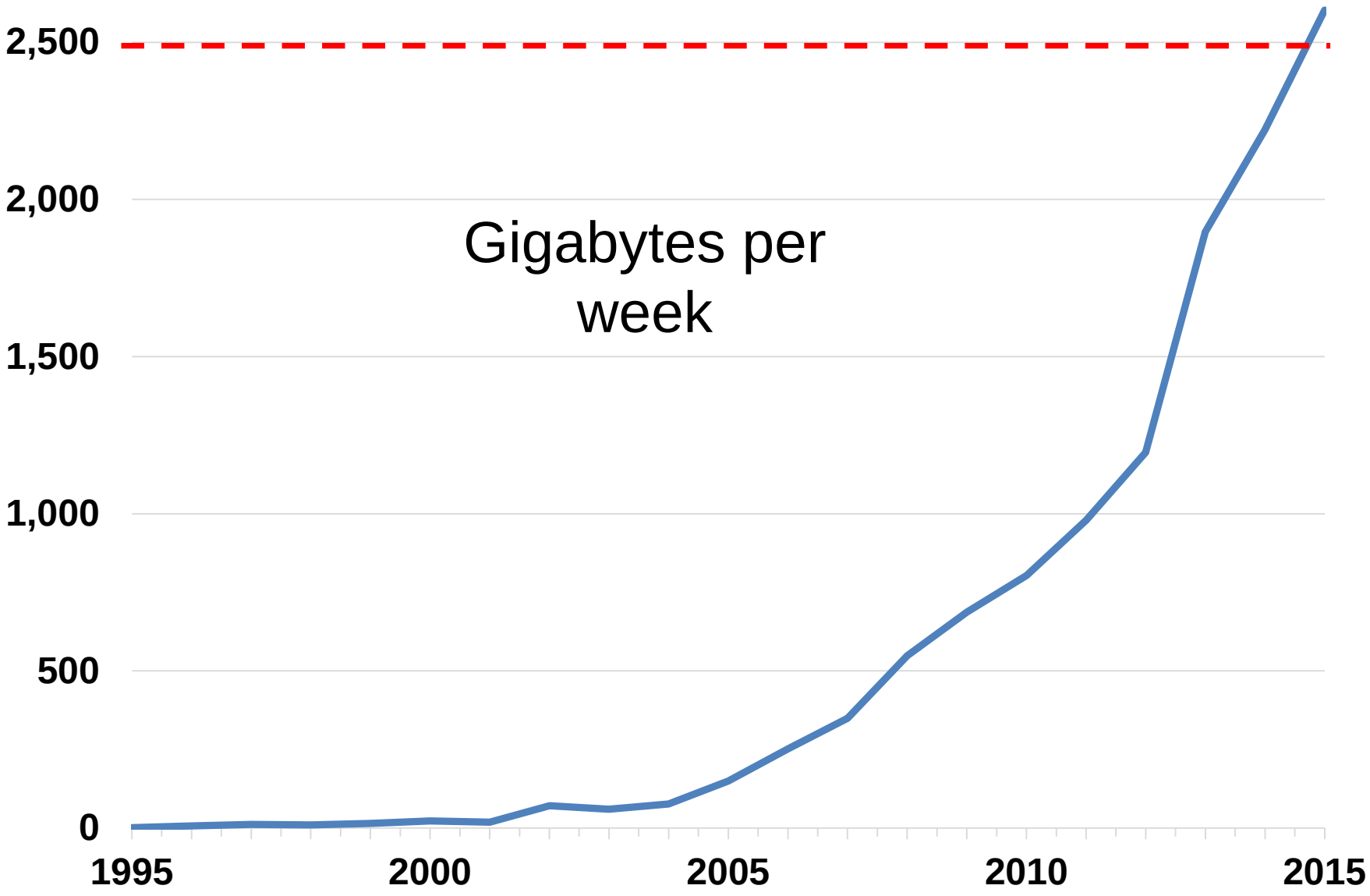
MPC

Minnesota Population Center
University of Minnesota

IPUMS Data Integration Projects

- World's largest archive of population data
  - Individual-level microdata describing ~3 billion persons enumerated in 100 countries
- 70,000 registered users from over 100 countries
- 1,500 publications annually
- Free to the research community

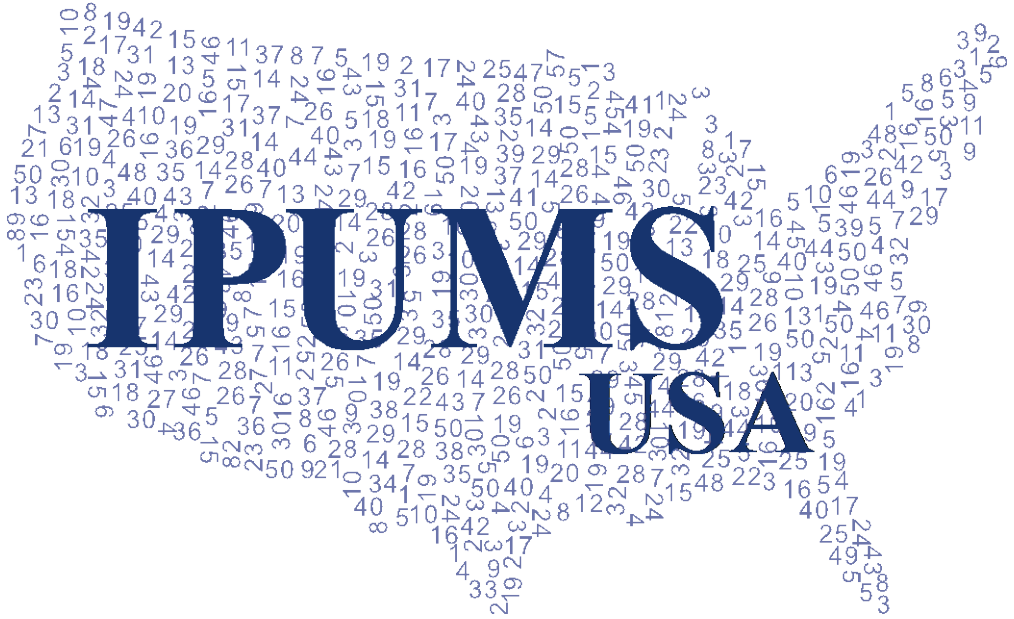
IPUMS Data Dissemination, 1995-2015



Outline

- International IPUMS projects
  - IPUMS-International
  - Integrated DHS
  - Terra Populus
- Spatial data integration
- Implications for SDGs


1960



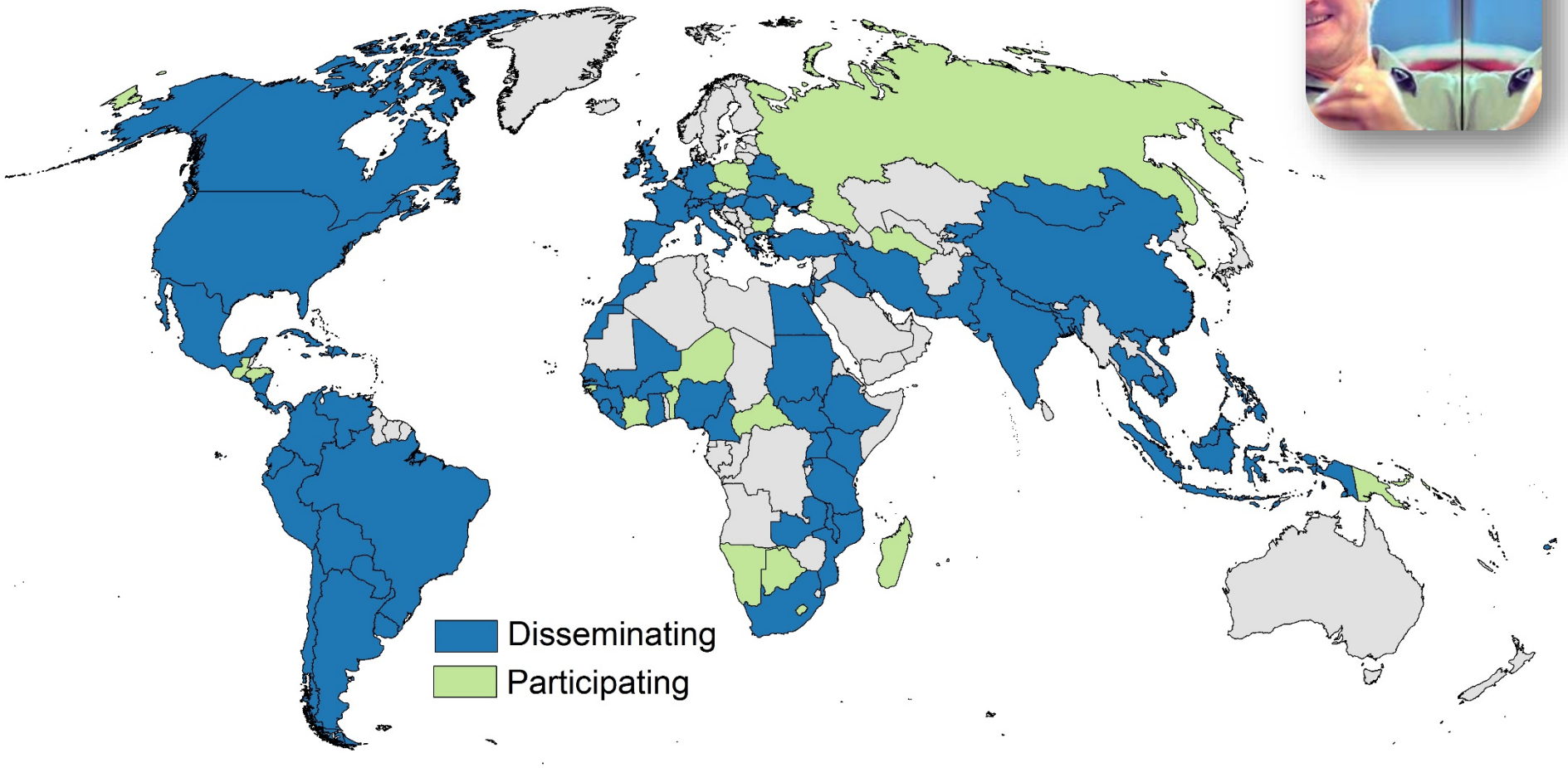
1991 IPUMS proposal: integrated data series for nine U.S. Censuses, 1850-1990

- Harmonized codes
- Consistent record layout
- Integrated documentation
- No loss of information

Preliminary version released in 1993, broad release in 1995

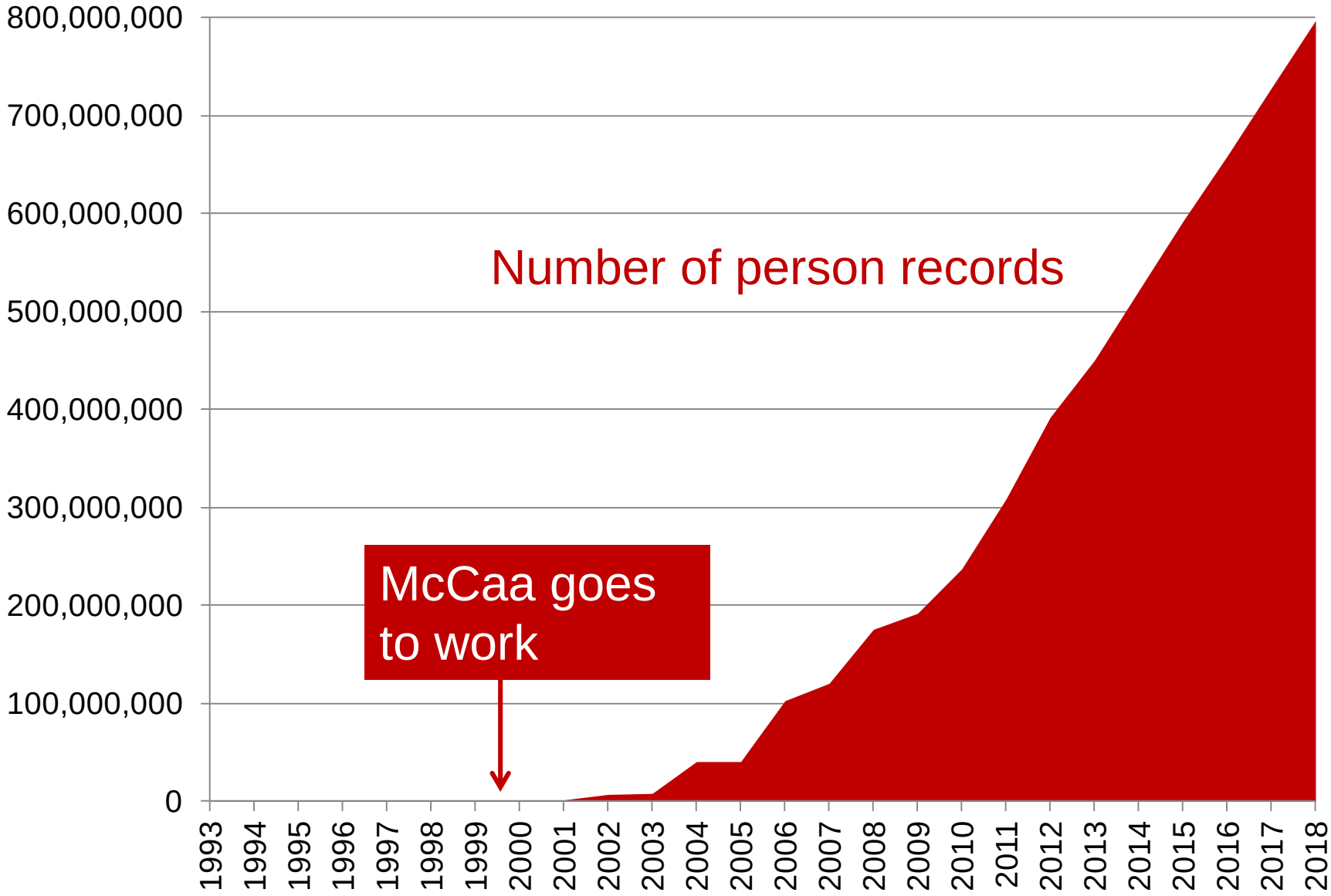
.

1999: IPUMS-International



Over 100 Collaborating National Statistical Agencies

Accessible Non-U.S. Census Microdata




Khartoum, CBS-Sudan

1973 Sudan Census tapes arrive


Dhaka, Bangladesh
Bureau of Statistics
COPY
CTL MRG OUTPUT
DIST-RAJSHAHI-80
-5984
CERTIFIED
6250 BPI


omputer·Link


Computer-Link

IPUMS Samples per Country

Argentina	5	Fiji	5	Malawi	3	Senegal	2
Armenia	1	France	7	Malaysia	4	Sierra Leone	1
Austria	4	Germany	4	Mali	2	Slovenia	1
Bangladesh	3	Ghana	2	Mexico	7	South Africa	3
Belarus	1	Greece	4	Mongolia	2	South Sudan	1
Bolivia	3	Guinea	2	Morocco	3	Spain	3
Brazil	6	Haiti	3	Nepal	1	Sudan	1
Burkina Faso	3	Hungary	4	Netherlands	3	Switzerland	4
Cambodia	2	India	5	Nicaragua	3	Tanzania	2
Cameroon	3	Indonesia	9	Nigeria	5	Thailand	4
Canada	4	Iran	1	Pakistan	3	Turkey	3
Chile	5	Iraq	1	Palestine	2	Uganda	2
China	2	Ireland	9	Panama	6	Ukraine	1
Colombia	4	Israel	1	Peru	2	UK	2
Costa Rica	4	Italy	1	Philippines	3	USA	7
Cuba	1	Jamaica	3	Portugal	3	Uruguay	6
Dominican Republic	5	Jordan	1	Puerto Rico	5	Venezuela	4
Ecuador	6	Kenya	5	Romania	3	Vietnam	3
Egypt	2	Kyrgyz Republic	2	Rwanda	2	Zambia	3
El Salvador	2	Liberia	2	Saint Lucia	2		

IPUMS Samples per Country

Argentina	5	Fiji	5	Malawi	3	Senegal	2
Armenia	1	France	7	Malaysia	4	Sierra Leone	1
Austria	4	Germany	4	Mali	2	Slovenia	1
Bangladesh	3	Ghana	2	Mexico	7	South Africa	3
Belarus	1	Greece	4	Mongolia	2	South Sudan	1
Bolivia	3	Guinea	2	Morocco	3	Spain	3
Brazil	6	Haiti	3	Nepal	1	Sudan	1
Burkina Faso	3	Hungary	4	Netherlands	3	Switzerland	4
Cambodia	2	India	5	Nicaragua	3	Tanzania	2
Cameroon	3	Indonesia	9	Nigeria	5	Thailand	4
Canada	4	Iran	1	Pakistan	3	Turkey	3
Chile	5	Iraq	1	Palestine	2	Uganda	2
China	2	Ireland	9	Panama	6	Ukraine	1
Colombia	4	Israel	1	Peru	2	UK	2
Costa Rica	4	Italy	1	Philippines	3	USA	7
Cuba	1	Jamaica	3	Portugal	3	Uruguay	6
Dominican Republic	5	Jordan	1	Puerto Rico	5	Venezuela	4
Ecuador	6	Kenya	5	Romania	3	Vietnam	3
Egypt	2	Kyrgyz Republic	2	Rwanda	2	Zambia	3
El Salvador	2	Liberia	2	Saint Lucia	2		

Many IPUMS variables are relevant to SDGs

- geographic location (places of 20,000+ persons in most samples, finer resolution in some cases)
- water supply, sewage, toilet, electricity, mobile telephones, Internet
- building materials—floor, roof, etc.
- educational attainment, literacy, school enrollment, disabilities
- economic activities, unemployment
- fertility history and child mortality

2011

IDHS

Integrated Demographic and Health Series

HOME | OVERVIEW | PARTNERS | STAF

Documentation

- Surveys
- MEASURE DHS Links
- Sample Descriptions
- Source Documents

Email Updates

Sign up to receive updates about IDHS

Welcome to IDHS

The Demographic and Health Surveys (DHS) are the main source of information on health
world. IDHS is designed to facilitate analysis of DHS data across time and space.

IDHS will:

- Focus initally on Africa and India
- Display variable availability across surveys
- Code data consistently across years and countries
- Provide variable-specific documentation
- Create customized multi-country and multi-year datasets
- Release data beginning in 2014

Why an Integrated DHS?

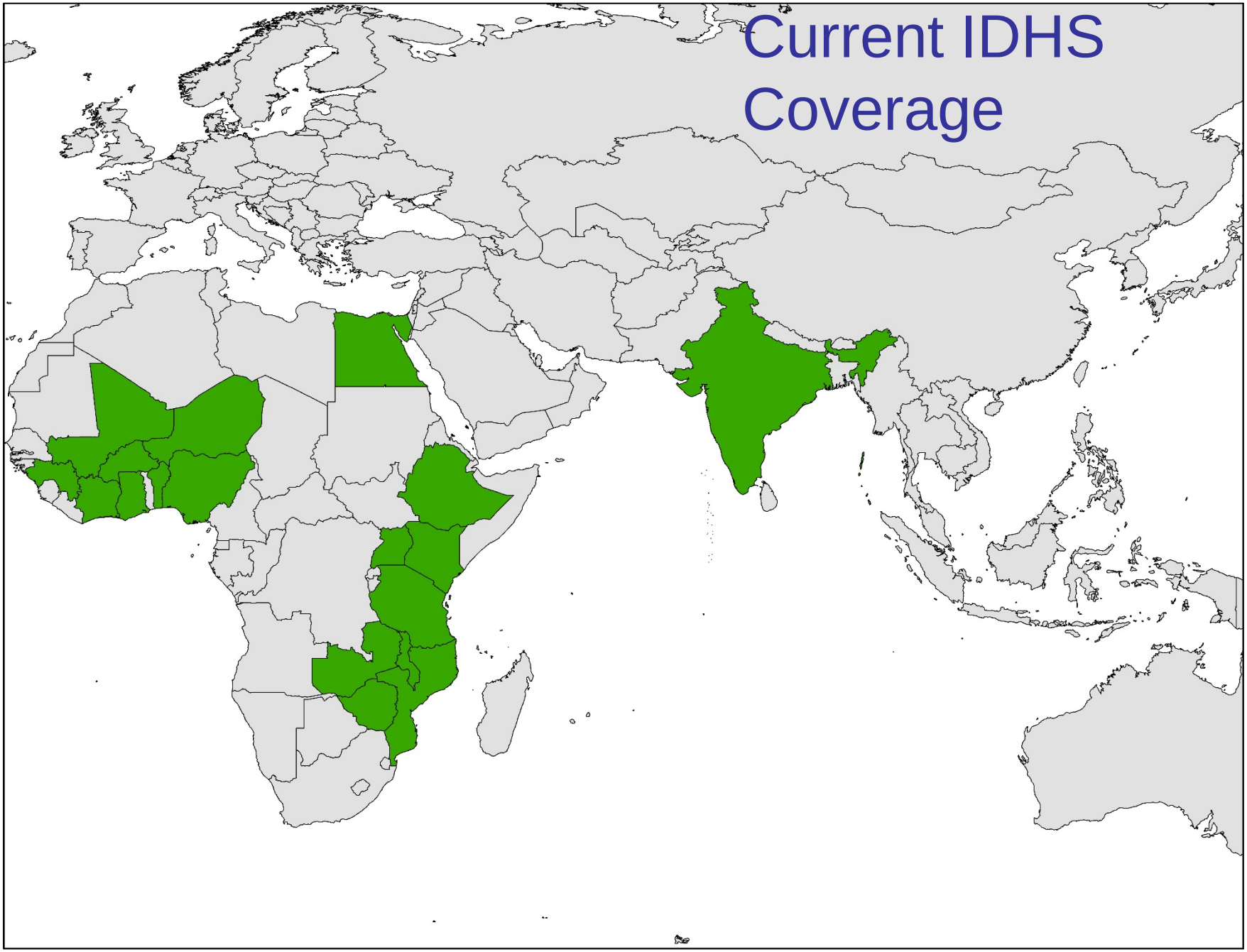
Motivation: DHS is incredibly valuable, but it's hard to capitalize on its full potential.

Problems:

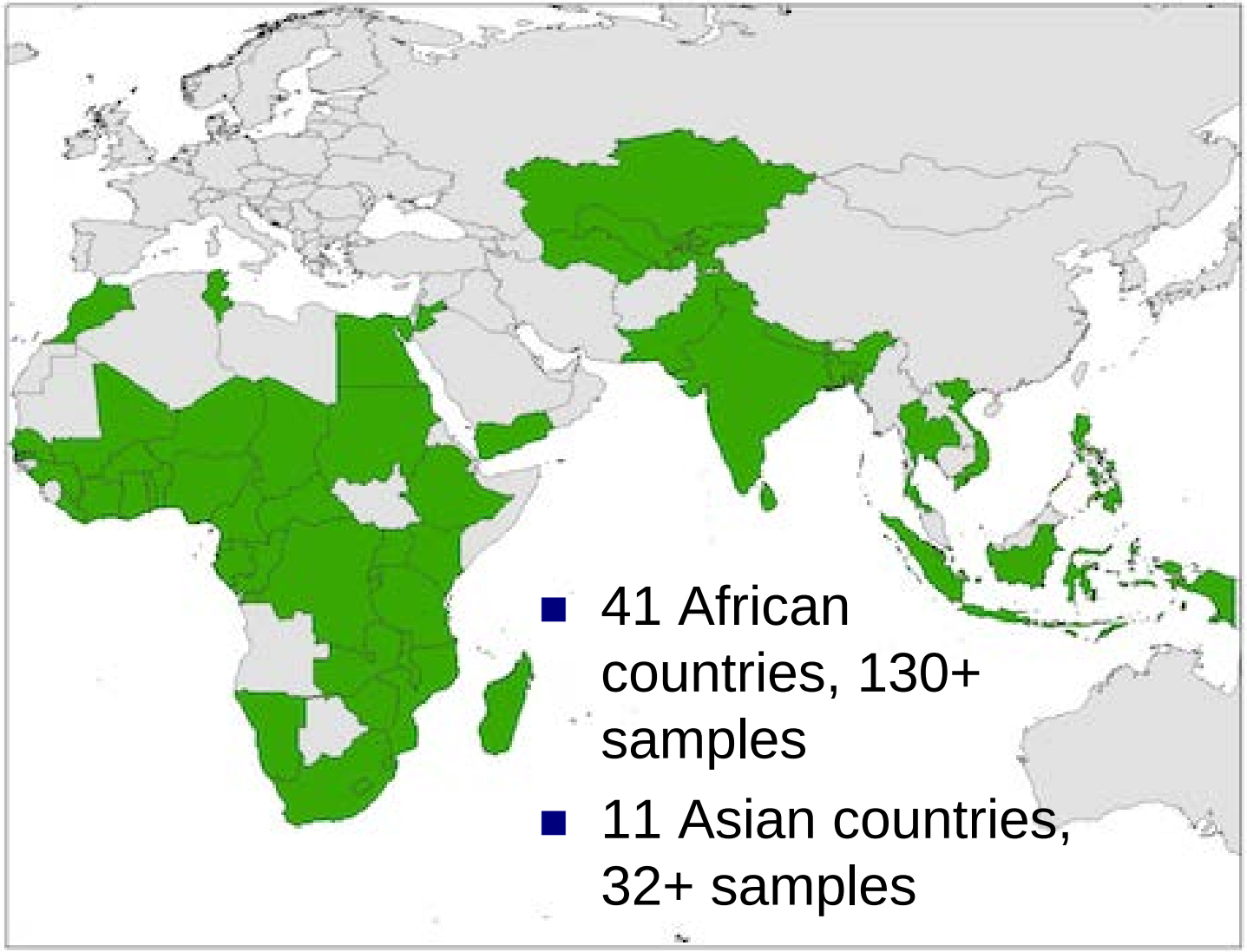
- Data discovery
- Dispersed documentation
- Data management
- Variable changes over time and between countries

Current IDHS Coverage

- 18 countries, 76 samples
- Women of childbearing age
- Children & birth files
- Men and couples files coming 2016


Current IDHS
Coverage

Next project phase (2016-2021)



TERRA POPULUS

Integrated Data on Population and Environment

HOME | OVERVIEW | DATA | PARTNERS | PEOPLE | CONTACT

Project Information

- Overview
- Data
  - Population Data
  - Environmental Data
  - Data Integration
- Partners
- People
  - Scientific Leadership
  - Executive Committee
  - Advisory Board
- Research Examples
- Presentations

Terra Populus: Integrated Data on Population and Environment

Terra Populus will integrate the world's population and environmental data, including

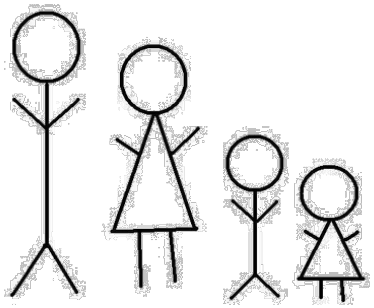
- Population censuses and surveys
- Land cover information from remote sensing
- Climate records from weather stations
- Land use records from statistical agencies

Data that are interoperable across time, space, and scientific domain will allow us to understand the dramatic transformation of the earth's inhabitants and their environment. TerraPop is developing this infrastructure to make it easier for researchers to use data describing people along with data describing the places they inhabit.

Breaking down disciplinary barriers

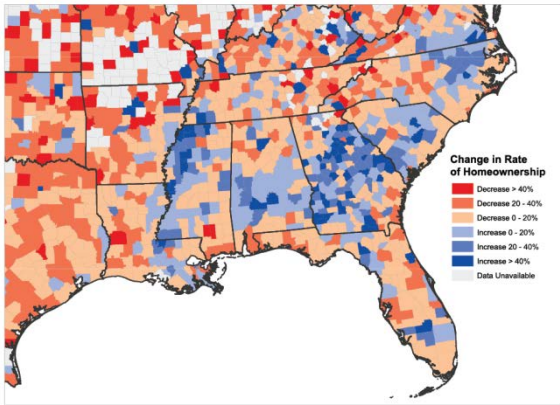
TerraPop integrates spatiotemporal data with different formats from different scientific domains to make them easily interoperable.

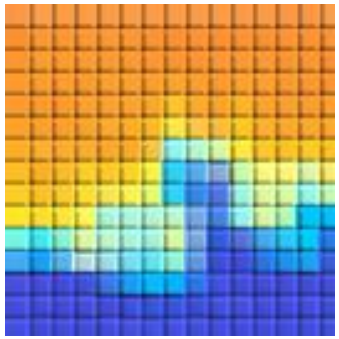
Three Source Data Formats



Microdata:
Characteristics of individuals and households

Small-area data:
Characteristics of places defined by administrative boundaries

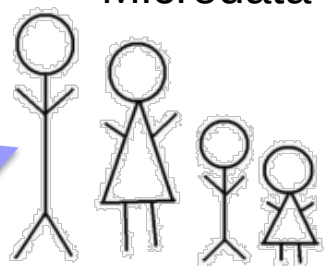




Raster data:
Values tied to spatial coordinates

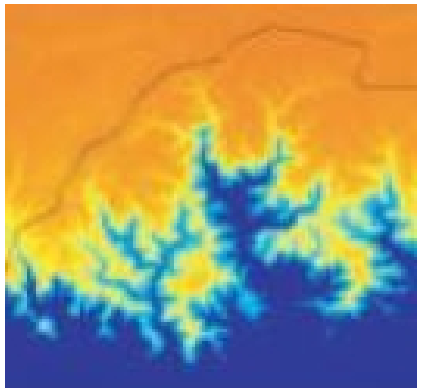
Location-Based Integration

Microdata

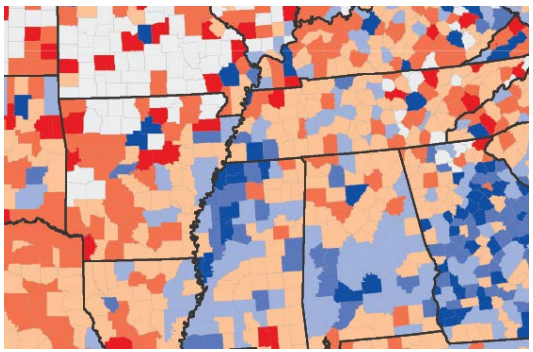


Mix and match variables originating in any of the data structures

Obtain output in the data structure most useful to you



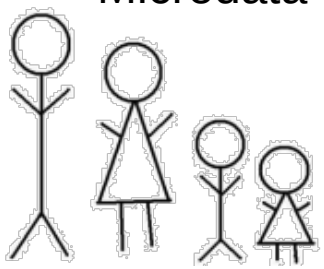
Rasters



Area-level data

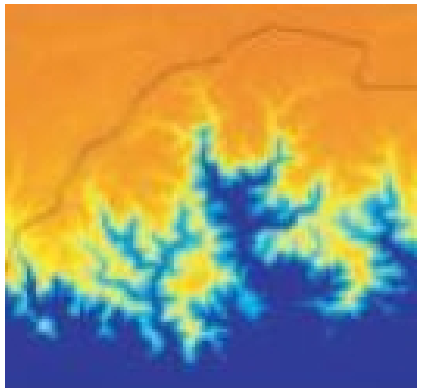
Location-Based Integration

Microdata

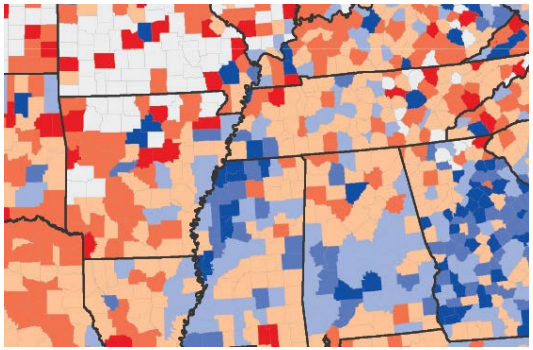


Individuals and households with their environmental and social context

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

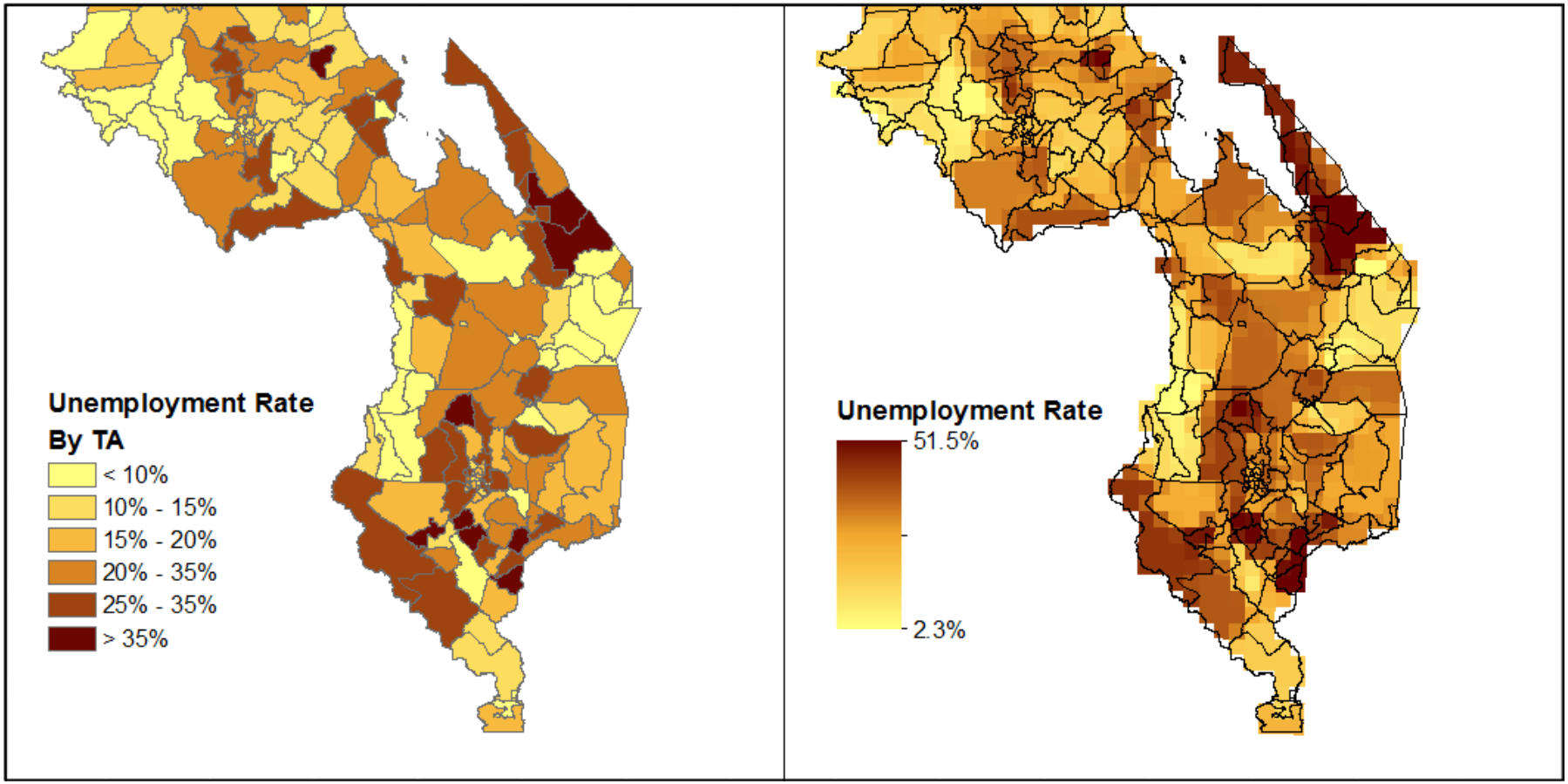


Area-level data

MPC

Minnesota Population Center
University of Minnesota

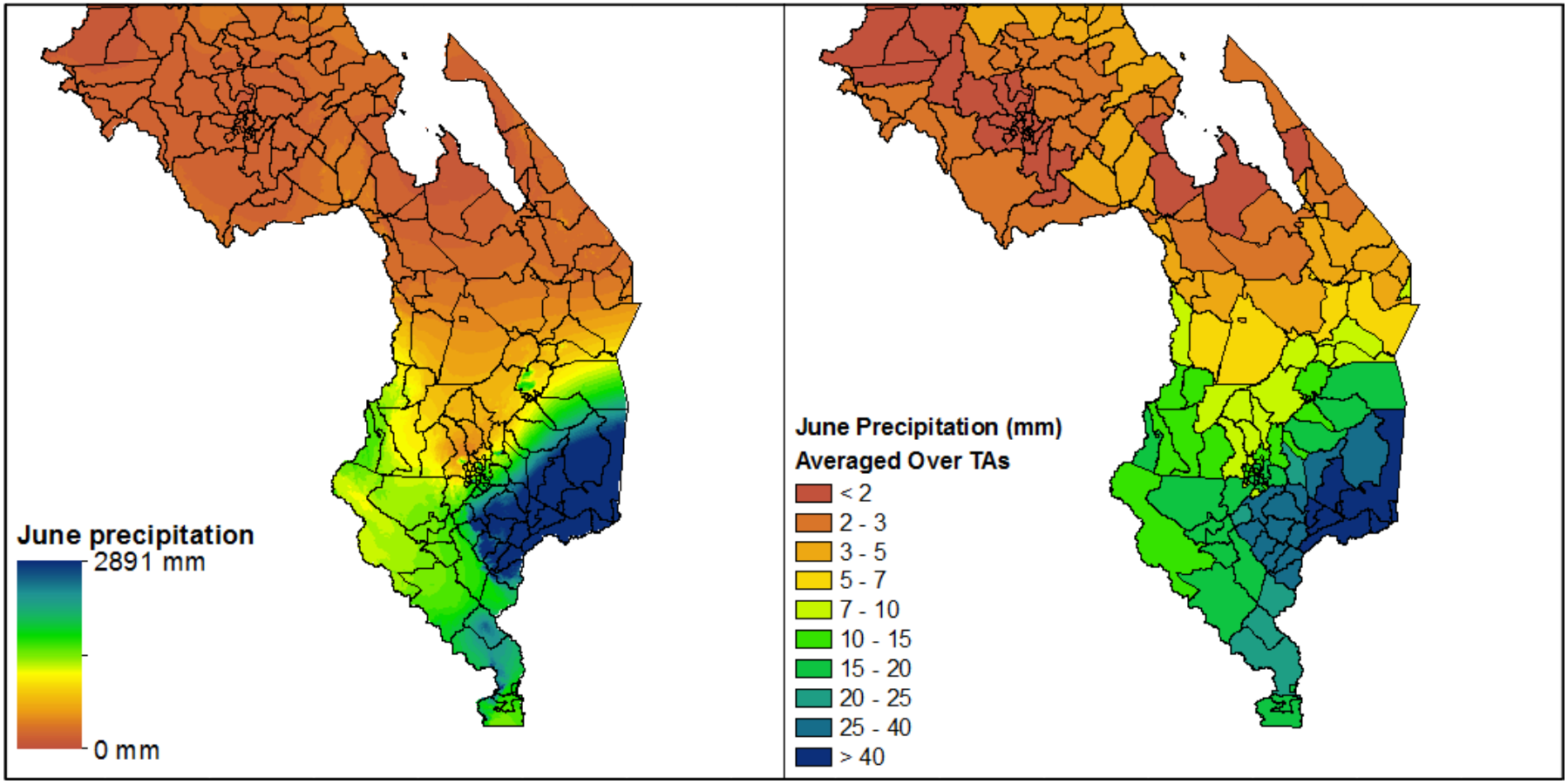
Rasterization of Area-Level Data



MPC

Minnesota Population Center
University of Minnesota

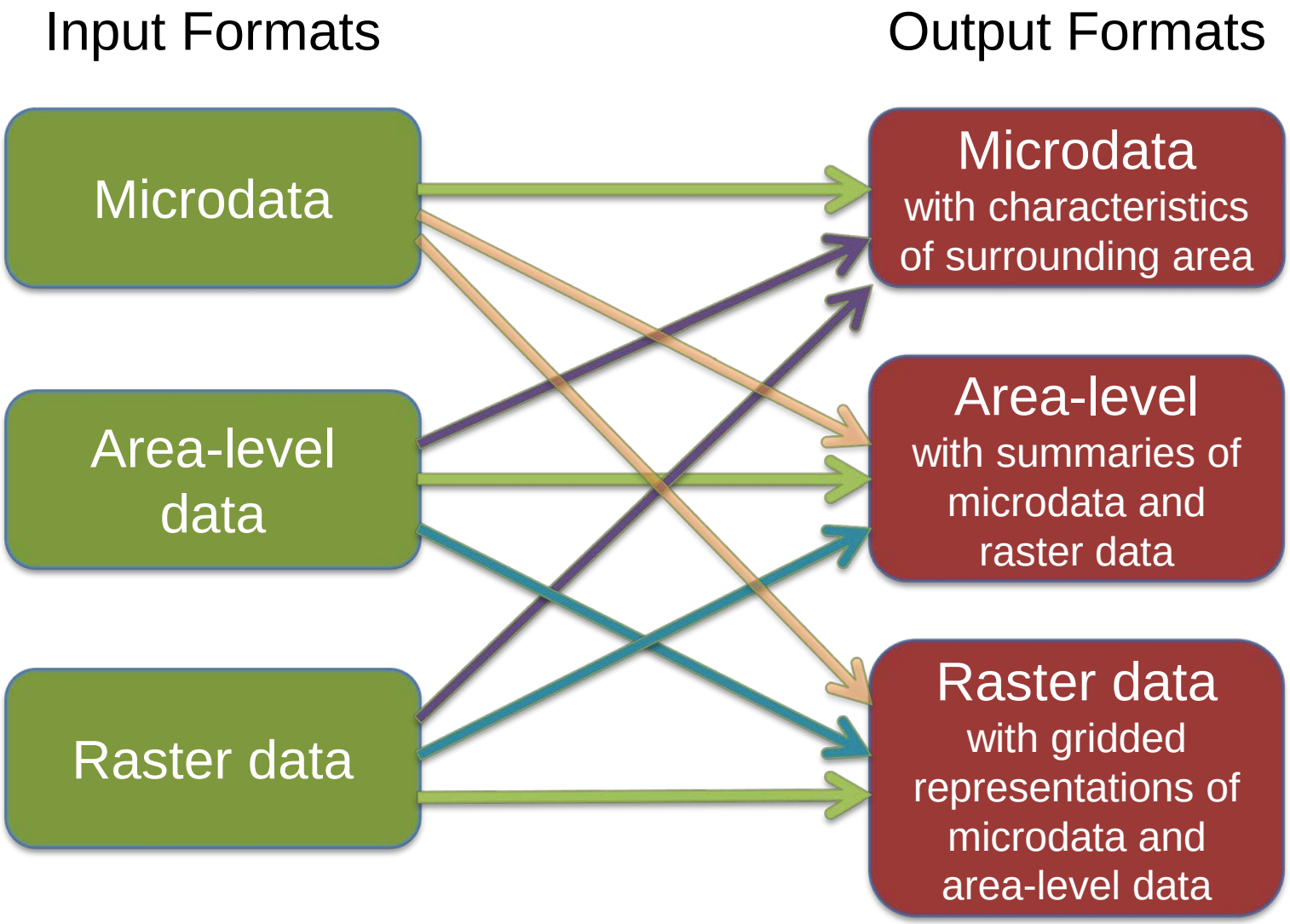
Area-Level Summary of Raster Data



MPC

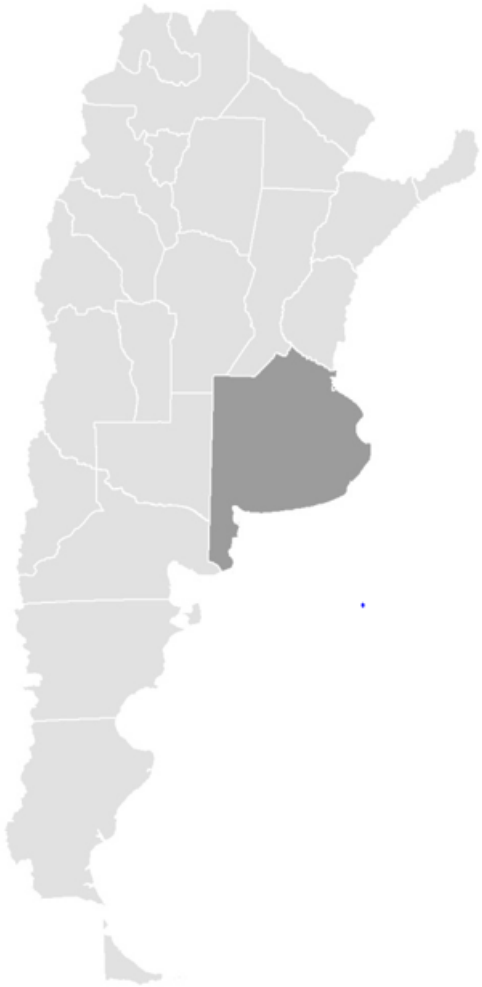
Minnesota Population Center
University of Minnesota

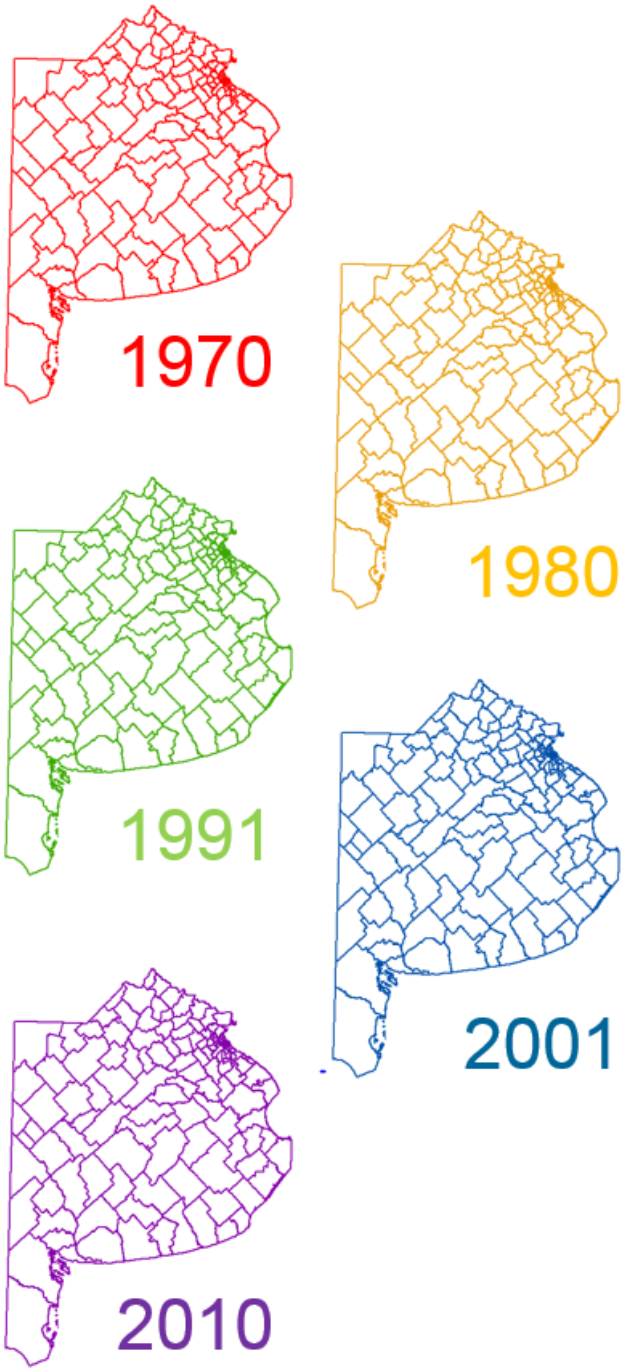
TerraPop Data Integration



Spatial data integration

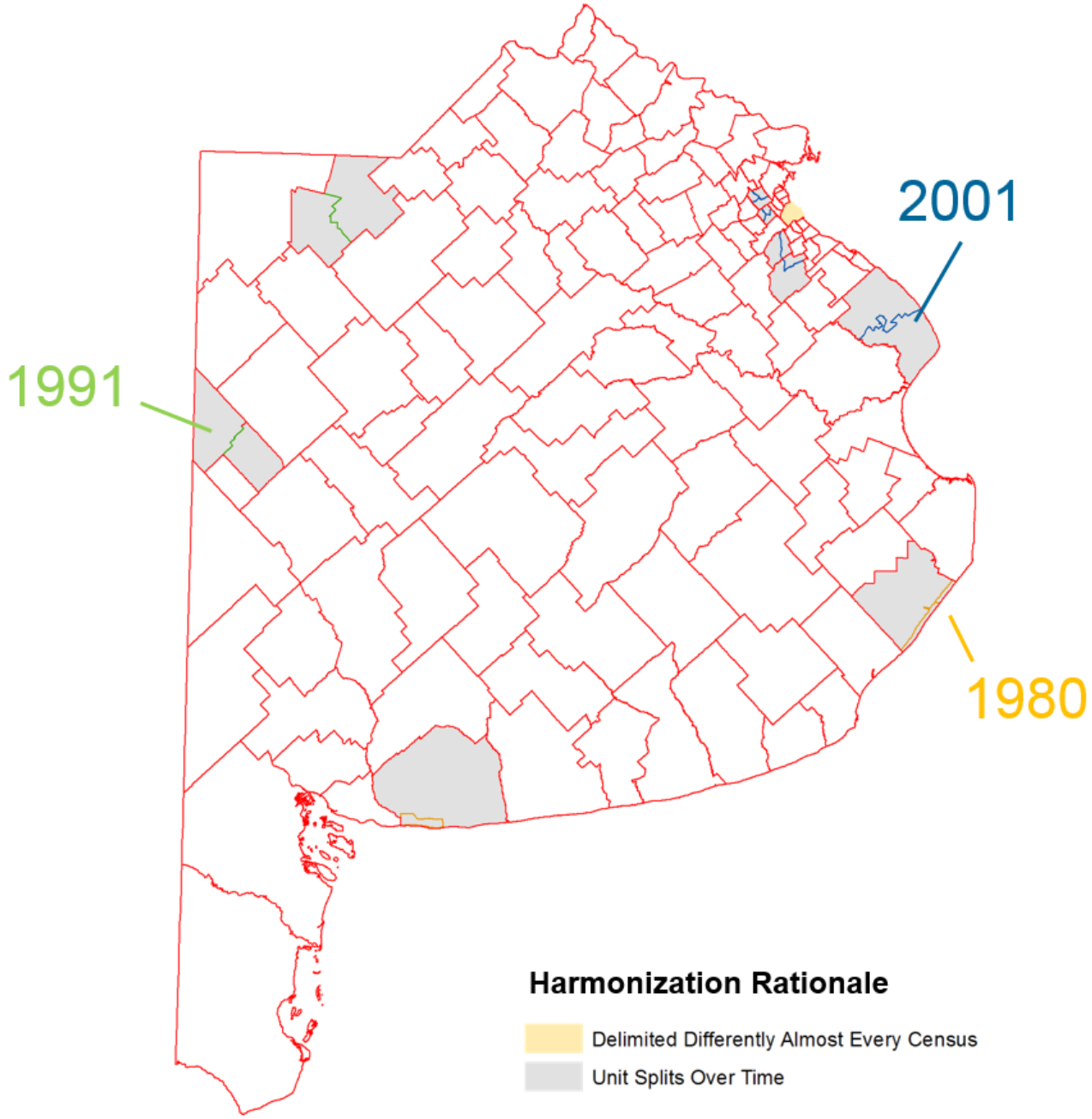
Buenos Aires province, Argentina





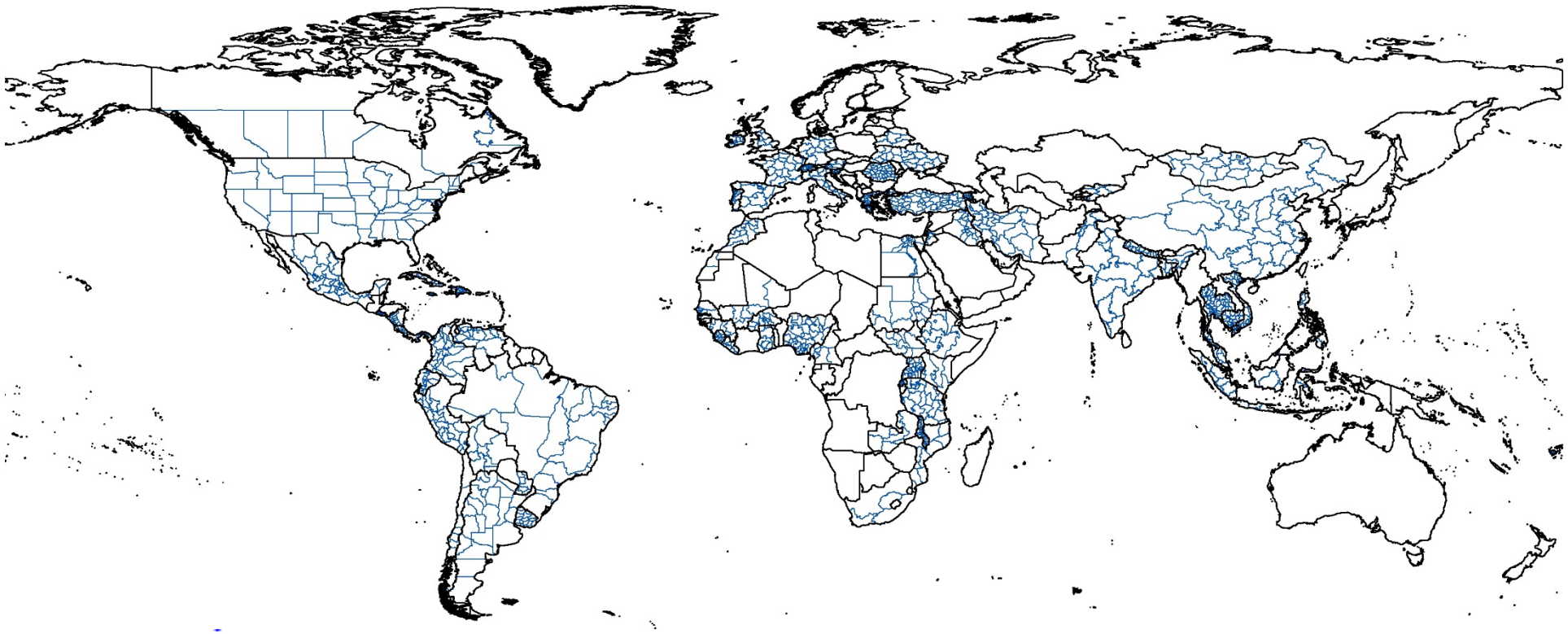
Spatial data integration

Some units must be merged or split to make footprint consistent over time



Progress on global harmonized boundaries

Integrated 1st-level boundaries for all countries released 2014

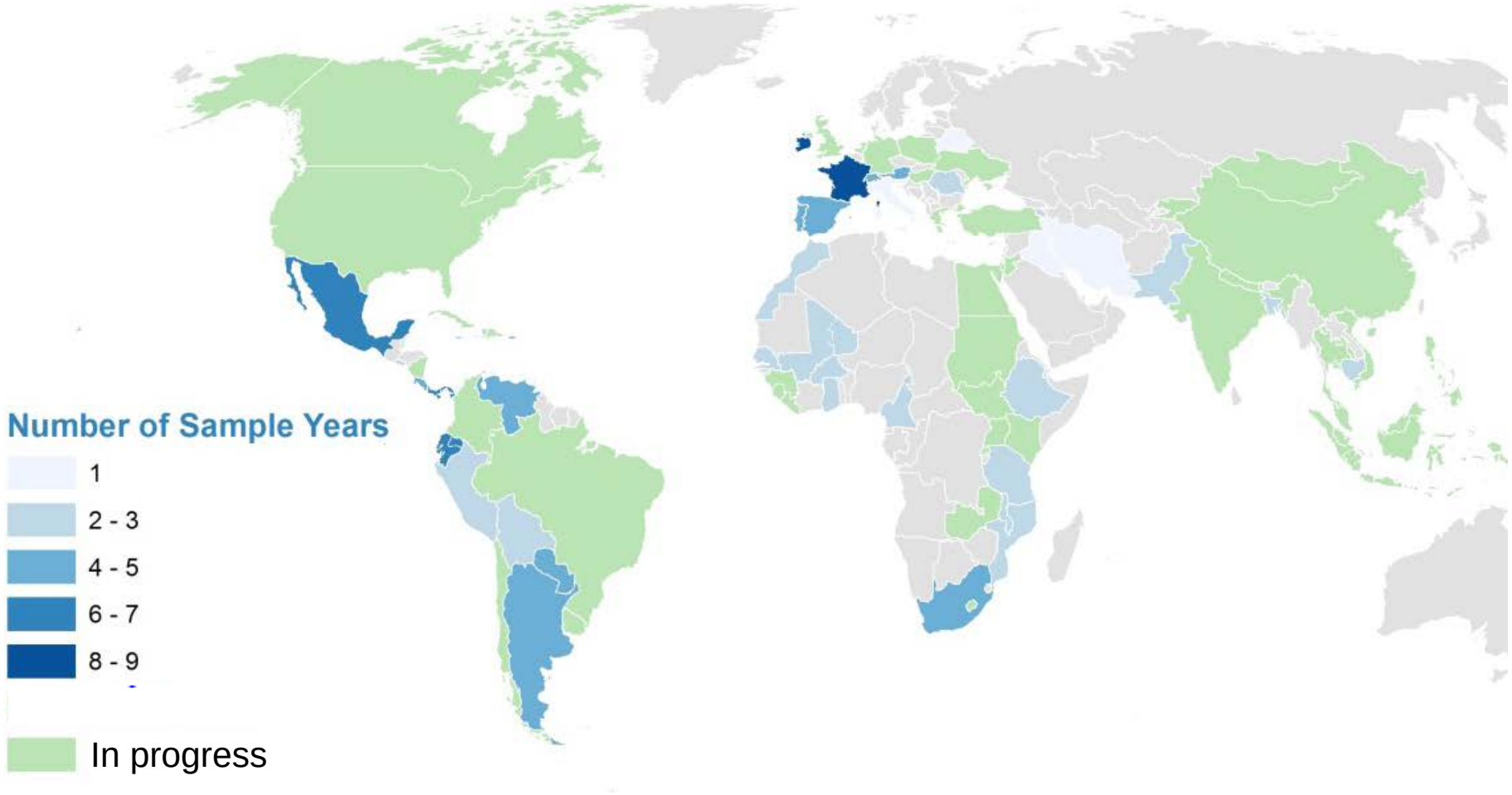


MPC

Minnesota Population Center
University of Minnesota

Progress on global harmonized boundaries

Integrated 2nd-level boundaries for 40 countries released last week



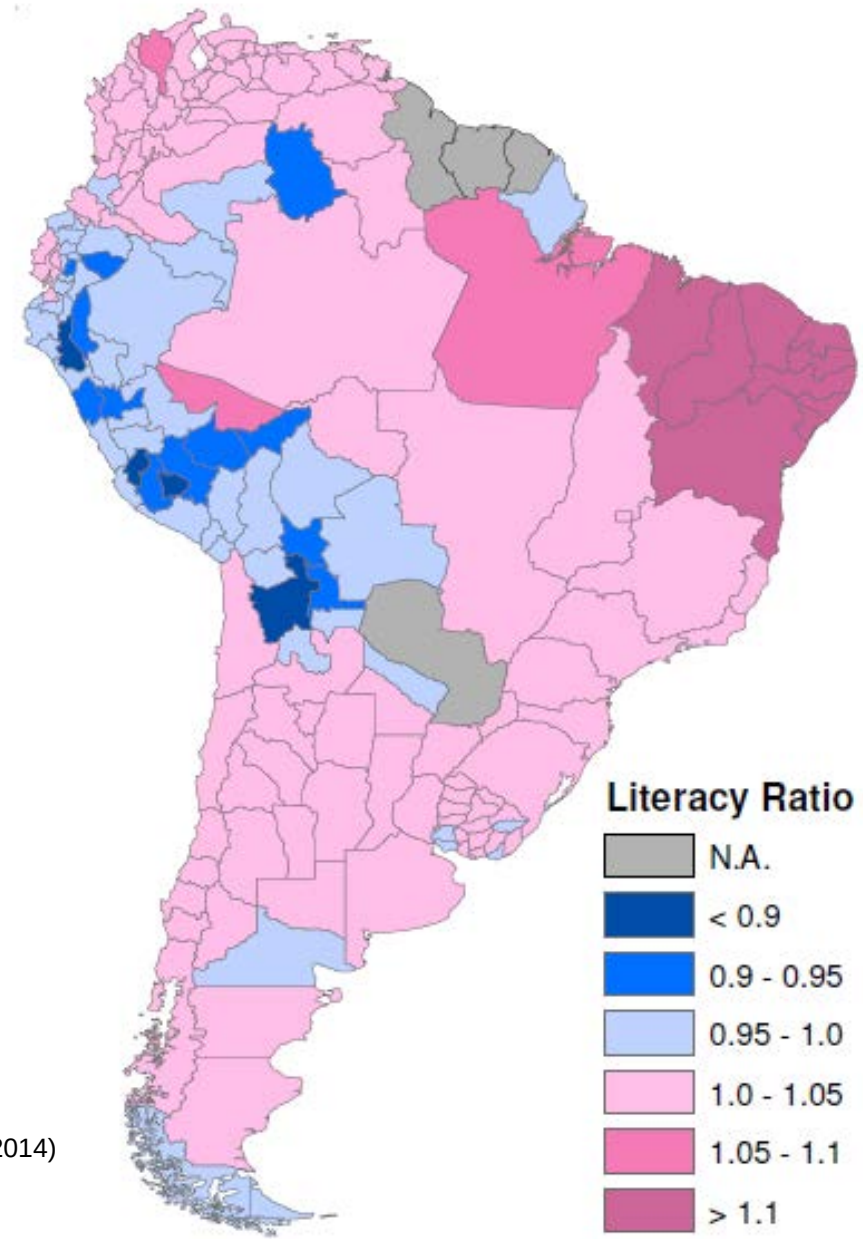
MPC

Minnesota Population Center
University of Minnesota

Millennium Development Goals

Ratio of literate women to men, 15-24 years old

1990 Census round

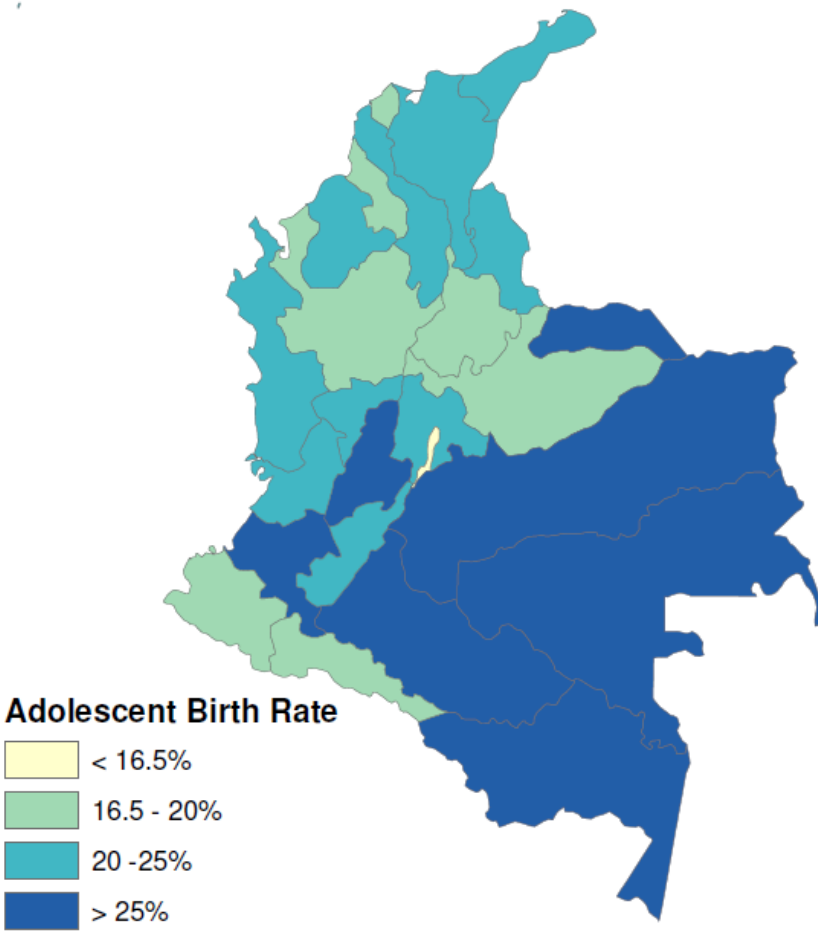


Source: Cuesta and Lovatón (2014)

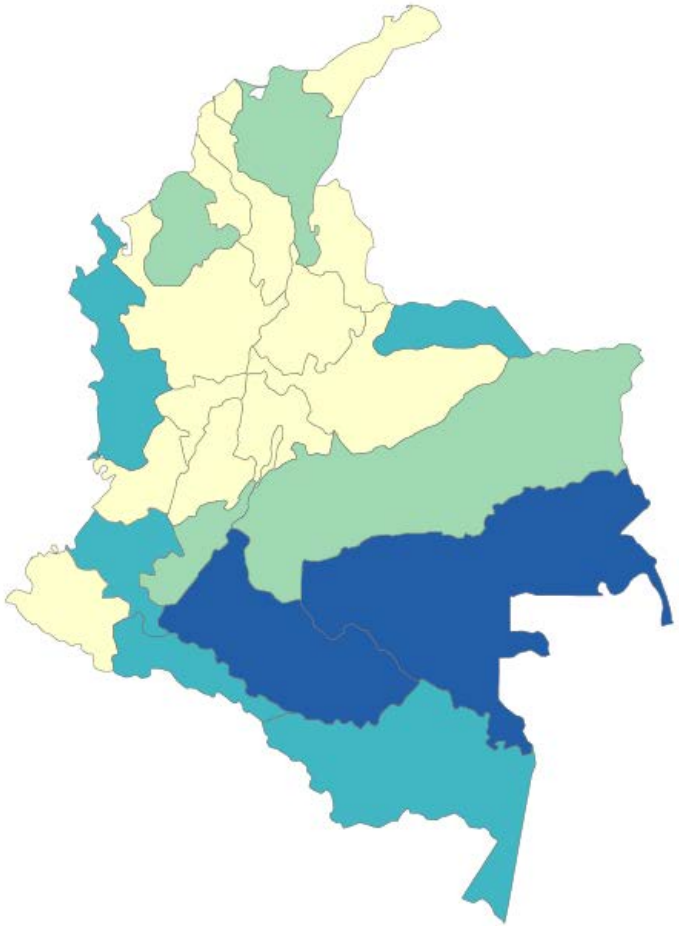
Millennium Development Goals

Colombia: Adolescent Birth Rate

Census 1993

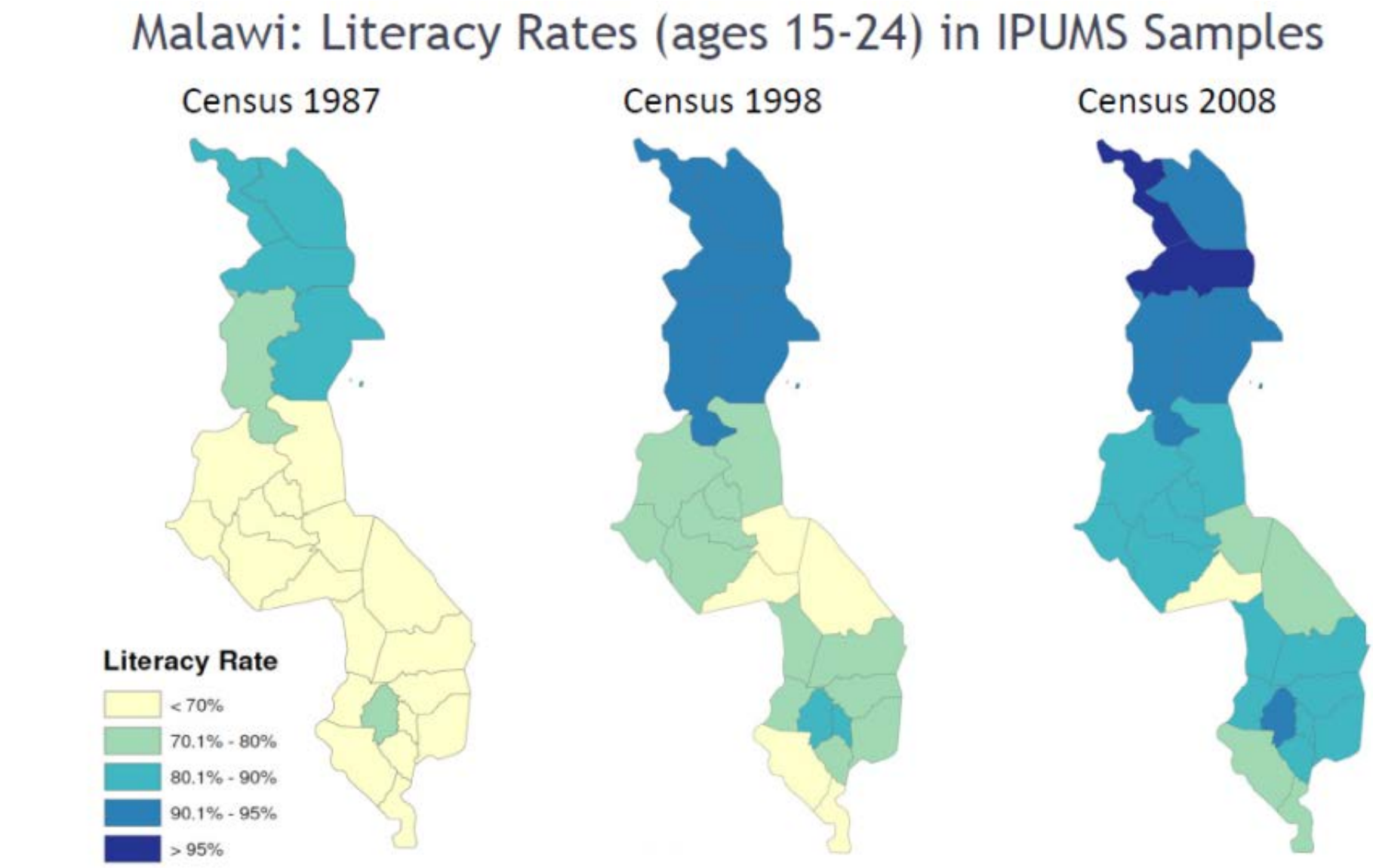


Census 2005



Source: Cuesta and Lovatón (2014)
Data Source: IPUMS-International, Minnesota Population Center

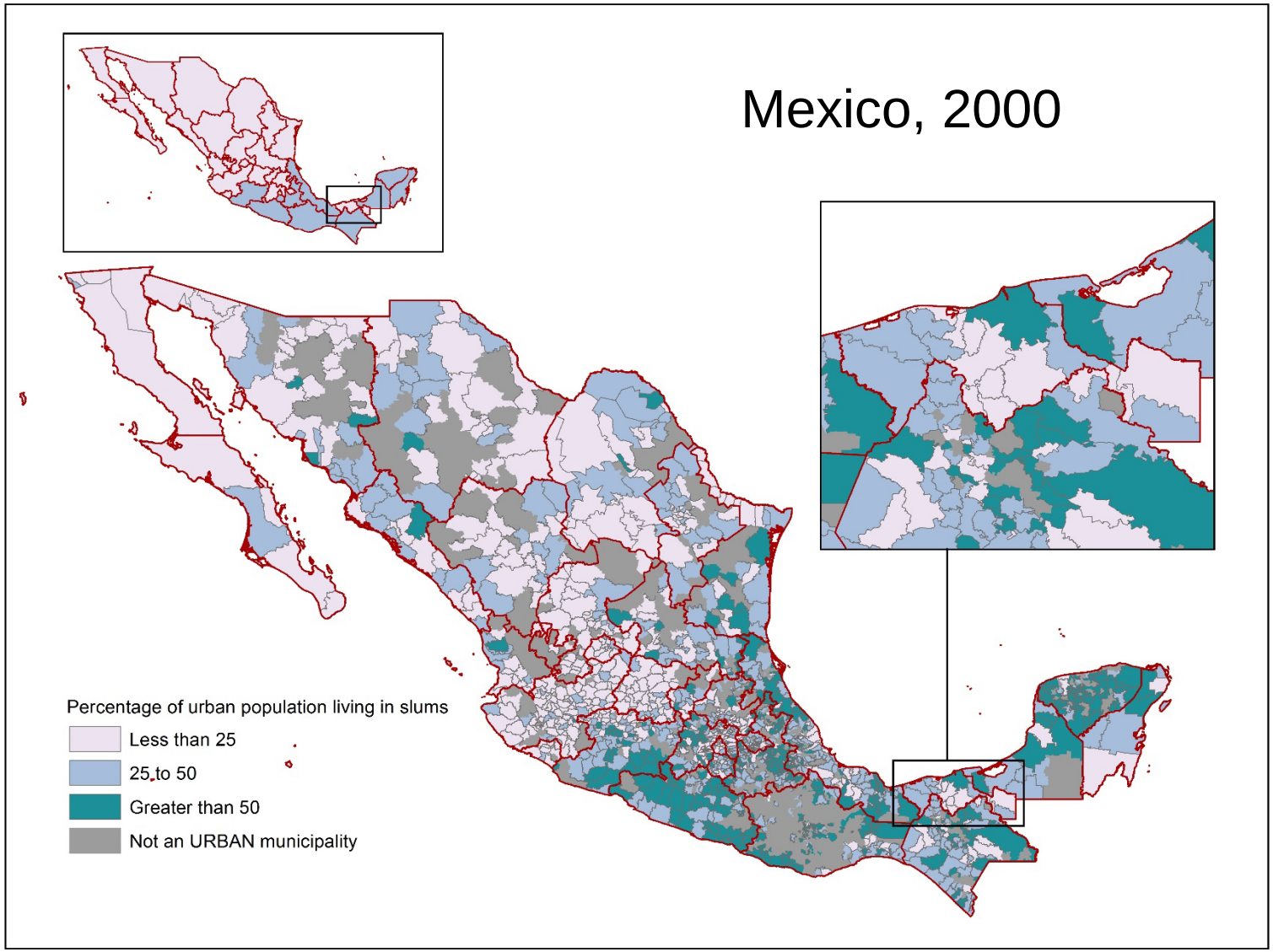
Millennium Development Goals



Data Source: IPUMS-International, Minnesota Population Center

Sustainable Development Goals

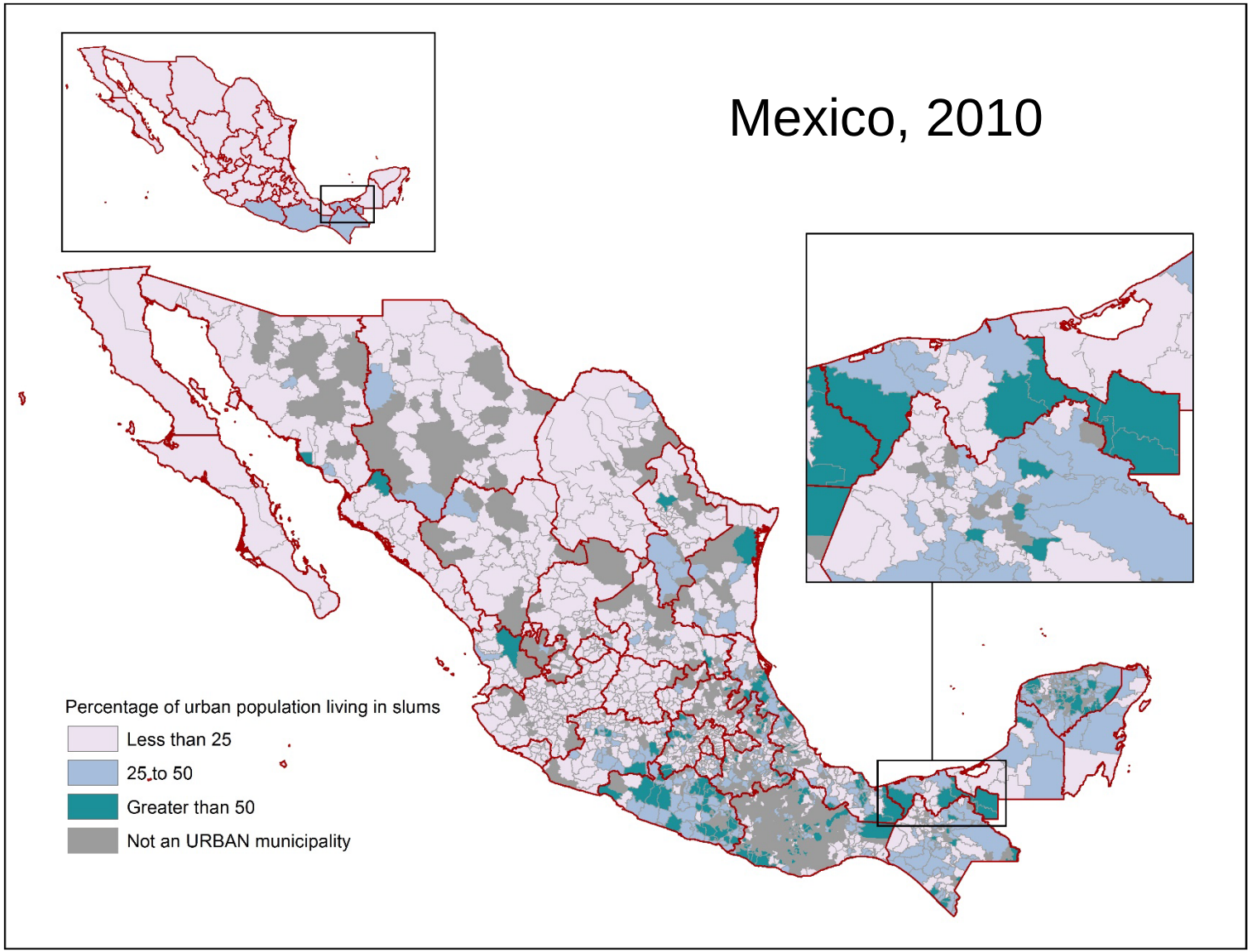
Percentage of the urban population living in slums



Source: IPUMS-International, Minnesota Population Center

Sustainable Development Goals

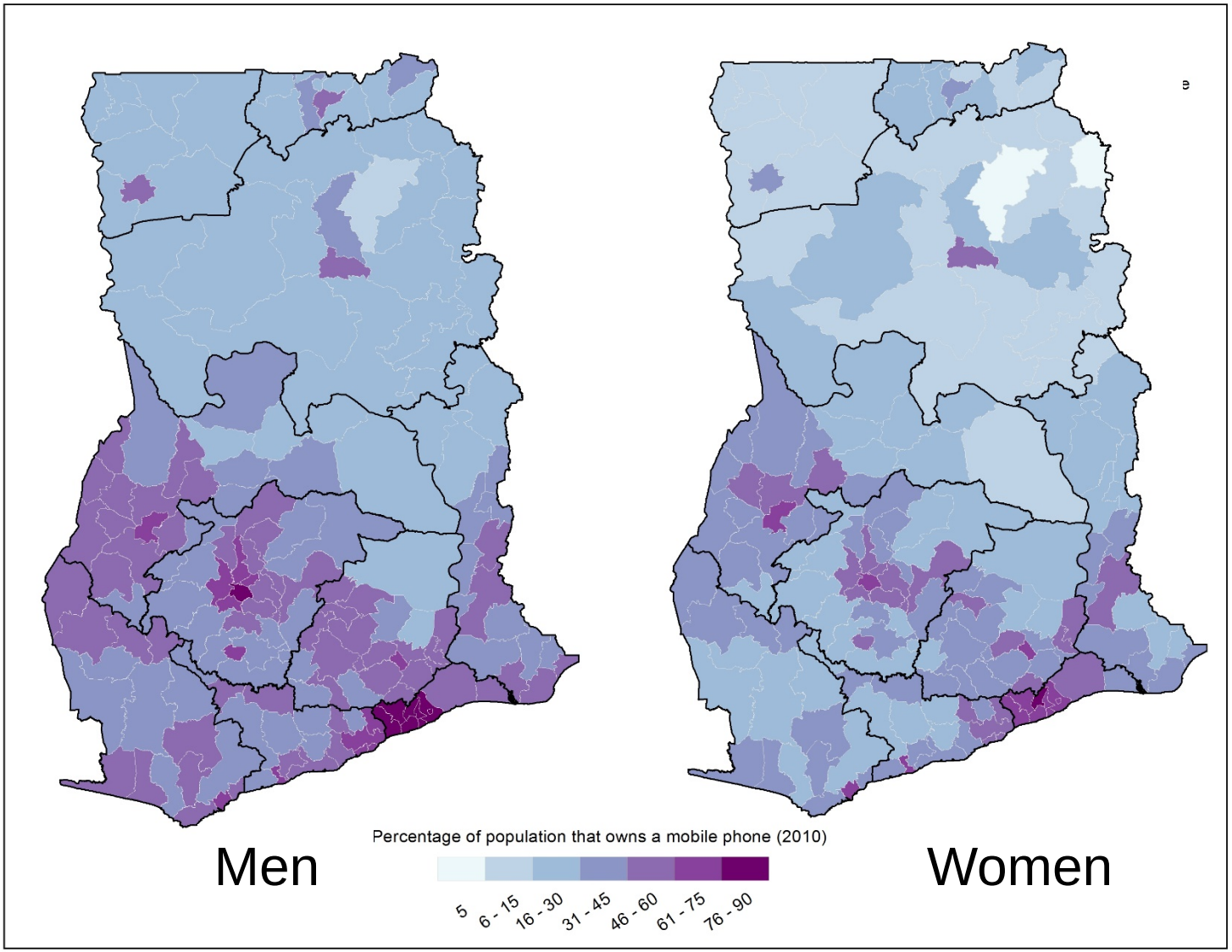
Percentage of the urban population living in slums



Source: IPUMS-International, Minnesota Population Center

Sustainable Development Goals

Percentage of the population that owns of mobile phone by sex, Ghana 2010



Conclusions

- Data integration is expensive, but it saves a lot of money in the long run, reduces the potential for error, and simplifies replication
- Administrative and survey data to be used for measurement of SDGs should be centrally integrated at the individual level wherever possible
- We need consistent geographic units over time to make sub-national estimates of change
- Sub-national estimates of change are essential for identifying places where progress has stalled and more resources are needed