

Conferencia “La Revolución Big Data en estudios sociales: ¿Qué  
hemos aprendido y hacia dónde vamos?”  
Colegio de México, 29 abril 2016

# Integración interdisciplinaria de fuentes de datos ‘tradicionales’ y Big Data en las ciencias sociales: desafíos y posibilidades

---

Susana B. Adamo

Center for International Earth Science Information Network (CIESIN)

The Earth Institute – Columbia University

[sadamo@ciesin.columbia.edu](mailto:sadamo@ciesin.columbia.edu)

# Organización de la presentación

- La revolución Big Data en ciencias sociales
- Los datos tradicionales
- Georreferenciación como marco de integración
- Beneficios de la integración de datos tradicionales y Big Data
- Desafíos adicionales
- Conclusiones

# La revolución Big Data en ciencias sociales

- ‘Explosión’ en la cantidad y diversidad de datos digitales de alta frecuencia en las ciencias sociales –telefonía móvil y registros de llamadas, banca electrónica, sensores remotos, etc. (Global Pulse 2012. *Big data for development: opportunities and challenges*).
  - Información proveniente de redes sociales (twitter, facebook, etc.) y todo producto de transacciones e interacciones (automáticas o no) digitalizadas
  - Orgánicos, no estructurados, masivos
  - Cuestiones de representatividad y sesgos desconocidos.

# La revolución Big Data en ciencias sociales (cont.)

Las “3 V”	Las “3 C”
Volumen Variedad Velocidad	Migajas (crumbs) Capacidades Comunidades de Big Data

Fuente: varias

Fuente: Harvard Humanitarian Initiative, MIT Media Lab and Overseas Development Institute. 2015. Opportunities and Requirements for Leveraging Big Data for Official Statistics and the Sustainable Development Goals in Latin America. **FULL DRAFT V1 FOR DISCUSSION.** Data-Pop Alliance

# Los datos “tradicionales”

- También llamados estructurados, de diseño, pequeños
  - Estadísticas oficiales (censos, encuestas de hogares, registros, estadísticas vitales)
  - Encuestas internacionales (DHS, MICS, etc.)
  - Bases de datos desarrolladas durante proyectos de investigación específicos
- Preguntas, capacidades, comunidades
- Menor resolución espacial y temporal
- Incremento notable en las últimas décadas:
  - Cantidad
  - Desagregación espacial
  - Armonización / interoperabilidad
  - Disponibilidad / Acceso

# Los datos tradicionales (cont.)

- “Big microdata”
  - La cantidad de micro datos disponibles, provenientes de censos y encuestas, ha crecido exponencialmente
  - Posibilita el estudio de procesos demográficos; y el testeado de modelos sociales y económicos  
(Ruggles, S. 2014. Big microdata for population research. *Demography*, 51:287-297)

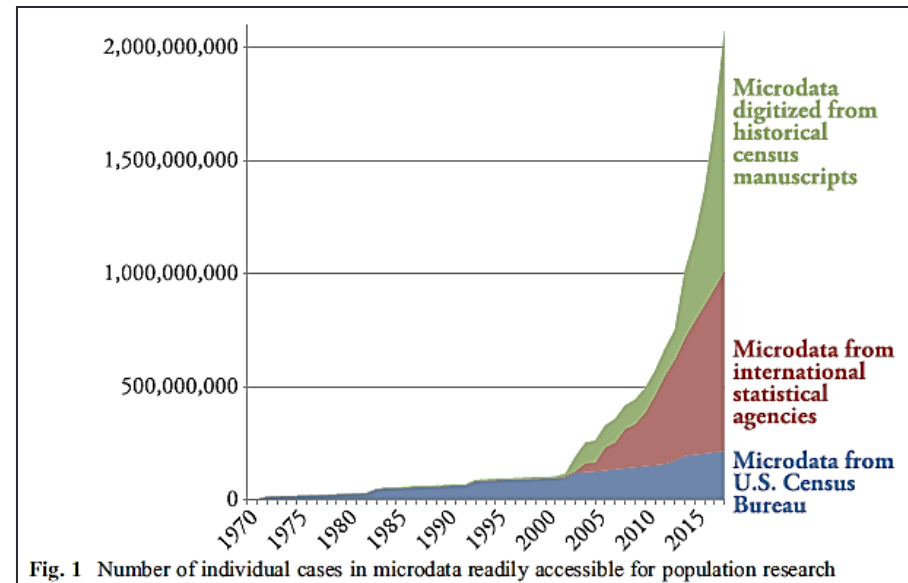


Fig. 1 Number of individual cases in microdata readily accessible for population research

Source: Ruggles 2014

# Fuentes tradicionales y Big Data

- Datos Tradicionales:
  - Contexto, marco, representatividad
  - Calidad y calibración

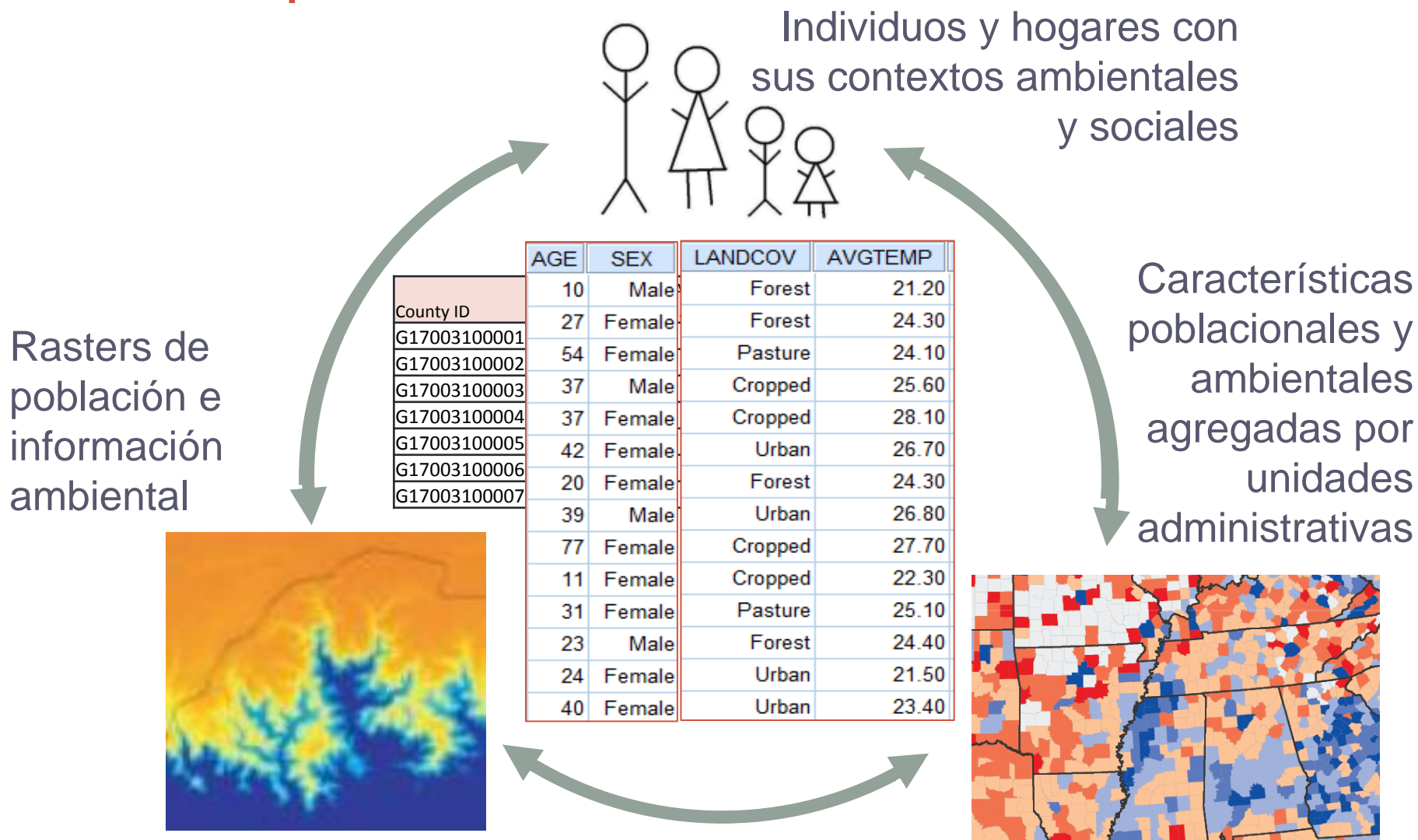
(IUSSP (2015). "The IUSSP on a Data Revolution for Development." Population and Development Review **41**(1): 172-177.)
- Big social data:
  - “Llenar los intersticios”: mayor desagregación espacial y mayor frecuencia, tiempo real, monitoreo
  - Variedad temática

(United Nations (2014). Prototype Global Sustainable Development Report. New York, United Nations Department of Economic and Social Affairs, Division for Sustainable Development.)
- Estudio de diferentes procesos sociales, o diferentes aspectos de un mismo proceso
- **Característica compartida: georreferenciación → integración basada en localización**

*The quality of organic data could may be lower than traditional data provides, but corrections can be made given a proper statistical understanding of the data (Keller, S., S. Koonin y S. Shipp. 2012. Big data and city living – what can it do for us? *Significance*, 9(4):4-7)*

*Organic data are voluminous but shallow: they often have no clearly defined universe, are unrepresentative of the general population, and do not systematically provide information about most of the things demographers care about, such as demographic behavior, education, work, and living conditions (Ruggles 2014)*

# Terra Populus

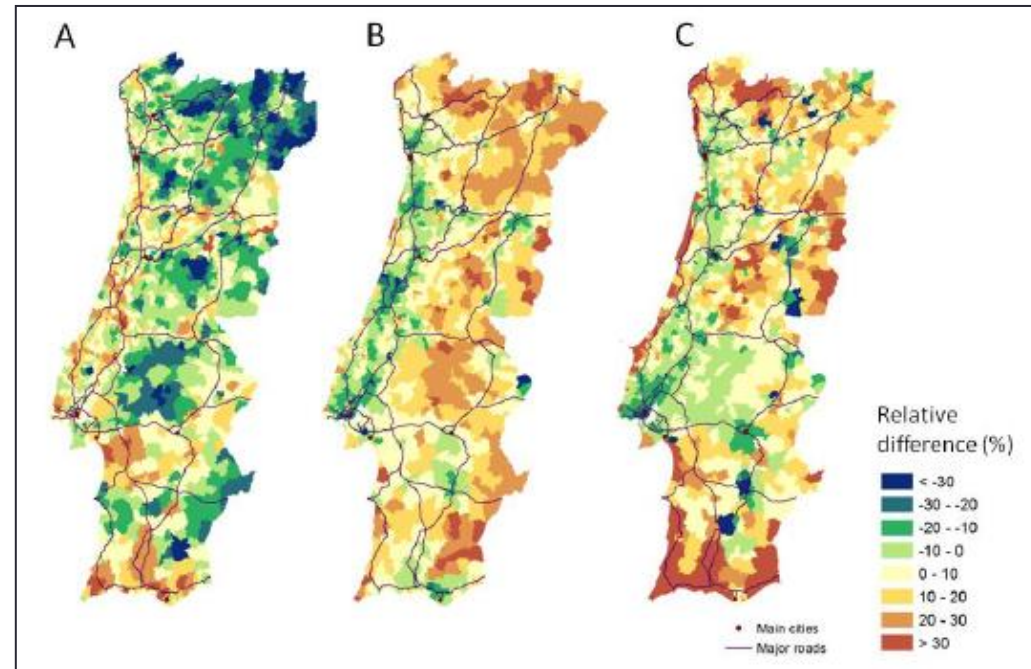


Versión beta disponible en <http://www.terrapop.org/>



# Estudio de la movilidad espacial con Big Data

- Registros de telefonía celular contribuyen al estudio de las “otras formas” de movilidad espacial:
  - Movimientos pendulares y desplazamientos cotidianos
  - Turismo
  - Migración laboral estacional
  - Desplazamientos en casos de desastre
- Densidad poblacional en tiempo cuasi real



Fuente: Deville, P., C. Linard, S. Martin, M. Gilbert, F. Stevens, A. Gaughan, V. Blondel and A. Tatem. 2014. Dynamic population mapping using mobile phone data. *PNAS* 2014 111 (45) 15888-15893

# Georreferenciación

- El giro espacial en ciencias sociales: considerar al espacio como una dimensión de análisis en la exploración y explicación de los procesos sociales
  - Espacio y lugar
  - Marcos contextuales espacio-temporales
- Dos elementos clave:
  - Acceso a información georreferenciada;
  - Desarrollo de nuevas herramientas para mapeo y análisis espacial: GPS, SIGs
  - Desarrollos de software libre y plataformas de procesamiento online

# Georreferenciación (cont.)

- Pensar críticamente sobre las propiedades de los datos espaciales:
  - Heterogeneidad, dependencia
  - Escala: extensión y resolución
  - Falacia ecológica y el problema de la unidad de área modificable
- La información georreferenciada ha crecido mas rápido que los métodos para analizarla, pero la brecha se esta cerrando aceleradamente;
- La resolución espacial de los sistema de adquisición de información puede ser diferente de la resolución adecuada para el análisis.

# Beneficios y necesidad de la integración

- **“La combinación de datos de diseño con datos orgánicos es la clave del futuro”**: la información mas efectiva para tener una visión clara de la sociedad incluirá video, sensores , imágenes satelitales, señales electrónicas, y datos demográficos y sociales tradicionales (Grooves, R. 2011. “Designed data” and “organic data”)
- Integración de datos para cubrir las necesidades de información del desarrollo sustentable
  - Integración de nuevas fuentes con los datos tradicionales para producir **información de alta calidad, más detallada, oportuna y relevante para muchos propósitos y usuarios**, especialmente para monitoreo;
  - El **aumento en la utilidad de los datos** a través de un mayor grado de apertura y transparencia, evitando la invasión a la privacidad y el abuso de los derechos humanos por el mal uso de la información sobre los individuos y grupos, y minimizando la desigualdad en la producción , el acceso y el uso de los datos (Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) (2014). *A world that counts: mobilizing the data revolution for sustainable development*, IEAG. page 6)

# Desafíos pendientes

- Interdisciplinariedad más compleja con incorporación de nuevos actores / disciplinas
- Brecha digital y reproducción de la desigualdad
- Acceso a los datos (tradicionales y Big Data)
- Cuestiones de privacidad y confidencialidad:
  - Efecto “localización”
  - Efecto “agregación”

# Para concluir

- Es necesario desarrollar más y mejores formas interdisciplinarias de manejo de datos (incluyendo Big Data)
  - Estructuras mixtas que combinan datos estructurados y no estructurados son un desafío, y requieren equipos con expertos en varias disciplinas incluyendo informática, data science y SIGs
- La georreferenciación constituye una plataforma viable para la integración de datos tradicionales y Big Data
  - *Data sources such as cell phone and Internet traffic, nighttime lights from satellite imagery, and even social networking content can be invaluable to social science if they were to be calibrated to specific populations and places. By combining big microdata with organic big data, we can enrich the microdata and frame the organic data (Ruggles 2014)*