

Seminario de Alto Nivel sobre Integración de fuentes de datos no tradicionales en los Sistemas Estadísticos Nacionales

*Panorama de los
diferentes tipos de
fuentes de datos no
tradicionales y su uso en
diferente dominios
estadísticos, tipos de
entidades que coleccionan y
producen dichas fuentes
de datos y su vinculación
con los Sistemas
Estadísticos Nacionales*

ECLAC/CEPAL, 1-2 October, 2018

Pascual Gerstenfeld

DIVISION DE ESTADISTICA

CEPAL

ORIGEN Y OBJETIVOS DE ESTE SEMINARIO DE ALTO NIVEL

- Dialogar para esclarecernos sobre los diferentes tipos de fuentes de datos no tradicionales, sus usos, y tipo de entidades que las recolectan así como su relación con los Sistemas Estadísticos Nacionales.
- Discernir las diferentes acepciones de datos tradicionales y no tradicionales en los diferentes dominios de estudio estadístico: económico, social, ambiental, y multi-dominios.
- Aclararnos las diferencias, límites y vínculos de este campo de actividad con los de otros campos como Big Data, Data Science, etc.
- La oportunidad y desafíos simultáneos que estimulan los ODS y la “Revolución de los Datos”
 - demanda por información que no puede ser atendida solamente por la vía de las fuentes habituales;
 - ni solamente por las ONEs o por una concepción restrictiva de los SEN.
- Fuentes de datos no tradicionales permiten:
 - *Ampliar temas*
 - *Ampliar desagregaciones*
 - *Disminuir costos*
 - *Estimular la integración de la información estadística y la geoespacial*

- Las fuentes de datos de las Estadísticas Oficiales para propósitos estadísticos deben ser tomados de todos los tipos de fuentes, sean encuestas, registros administrativos u otras fuentes.

- Las Oficinas de Estadística deben elegir la o las fuentes de acuerdo con la calidad, temporalidad, costos y carga de trabajo que signifique para los respondientes.

QUE DICEN LOS PRINCIPIOS FUNDAMENTALES DE LAS ESTADÍSTICAS OFICIALES SOBRE ESTE TOPICO

Principios Generales para integrar distintas fuentes de datos en los Sistemas Estadísticos Nacionales



**Que dice el
Cape Town
Global Action
Plan sobre
Datos para el
Desarrollo
Sostenible**

Objetivo 2.3: Facilitar la aplicación de nuevas tecnologías y nuevas fuentes de datos en las actividades de los SEN

- **Identificar y remover barreras para el uso de nuevas fuentes de datos, incluyendo datos de registros administrativos y datos de fuentes innovativas y coordinar esfuerzos para incorporarlos en los programas estadísticos de los SEN (...).**
- **Desarrollar guías para el uso de nuevos e innovativos datos generados fuera del Sistema Estadístico Oficial, para incorporarlos en las**

Objetivo 3.1: Aumentar la integración de diferentes fuentes de datos: censos, registros, nuevas fuentes.

Objetivo 3.2: Mejorar los registros estadísticos nacionales y expandir el uso de registros administrativos integrándolos con datos de encuestas y otras fuentes de datos (...).

Objetivo 3.4: Integrar los programas de producción de información geoespacial y estadística en todos los niveles.

Grupo Global de trabajo sobre Big Data para las Estadísticas Oficiales



Este Grupo de trabajo fue establecido por la StatComm para investigar los beneficios y desafíos que implica el uso de Big Data, particularmente para el seguimiento de los ODS.



Trabajar sobre topics relacionados con: metodología, calidad, tecnología, legislación sobre acceso a datos y privacidad, gestión de datos y financiamiento.



Grupos de Tareas

Datos de imágenes satelitales

Datos de teléfonos móviles

Datos de medios de comunicación social

Datos de scanner

Habilidades, entrenamiento y creación de capacidades



Proponer una plataforma global para datos confiables, servicios y aplicaciones sobre un base de “data-hubs” y una arquitectura federal.

A QUE NOS REFERIMOS CON FUENTES NO TRADICIONALES: sus desafíos y potencialidades

Registros administrativos → son una fuente de datos tradicional

(ej. Registro civil), pero hay un amplio espacio para su mejor aprovechamiento estadístico.

- a. Acceso a registros de diversos organismos, con posibilidad de vincularlos mediante identificador único, abre posibilidades.**
- b. Ej: registros de seguridad social, educación, impuestos, beneficiarios de programas sociales.**

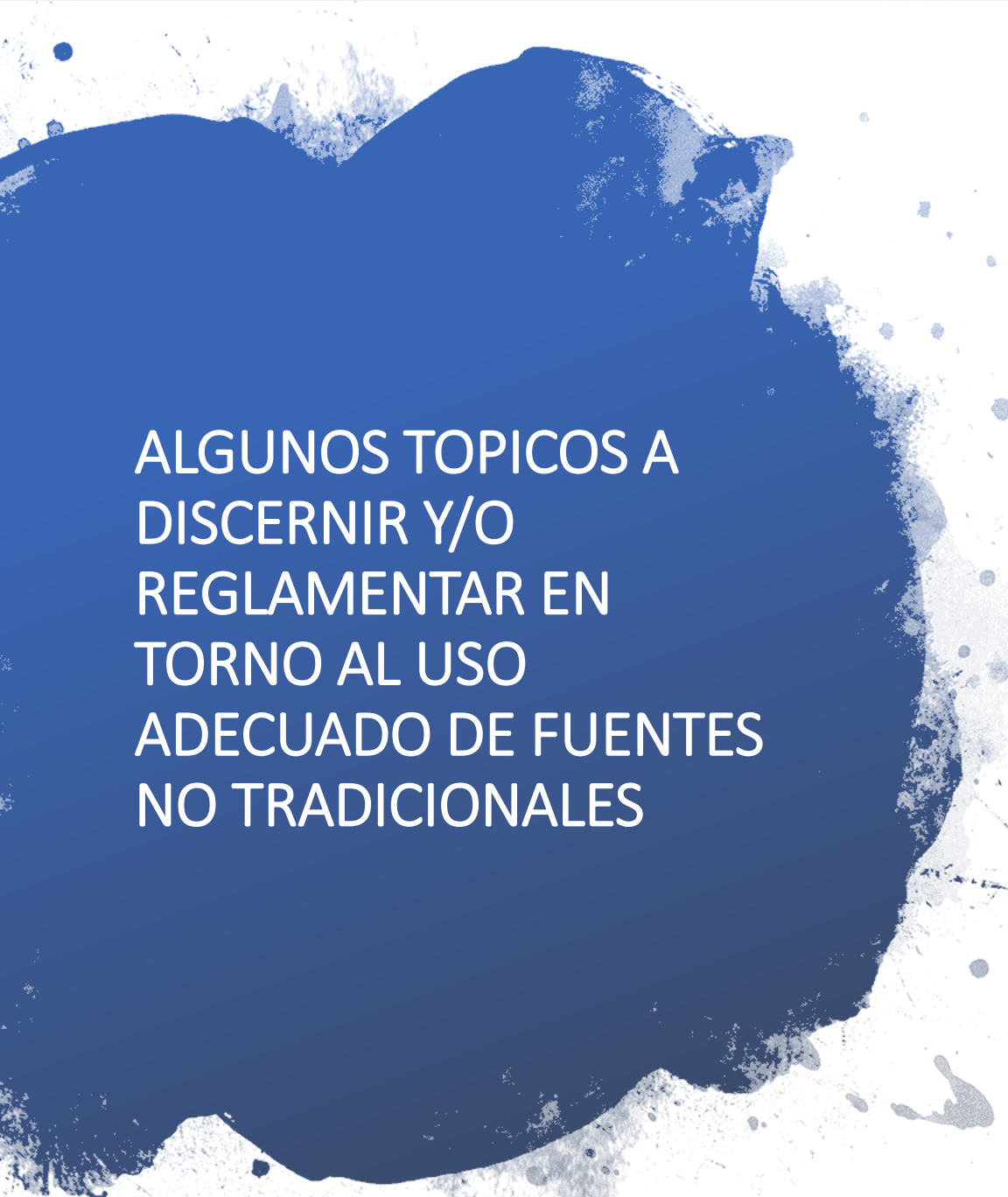
• Imágenes satelitales → Ya se usan para estadísticas ambientales. Pero se busca un aprovechamiento mediante vínculo entre geoespacialidad y estadística + producción de nuevas estadísticas con base en información de imágenes satelitales

• Big data digitalizada (datos de teléfonos móviles, transacciones comerciales, bases de datos de transacciones financieras)

• Datos producidos por otros actores tradicionalmente fuera del SEN (ej. Universidades, Centros de investigación)

DESAFIOS PARTICULARES QUE PRESENTAN ESTAS FUENTES

- **Desafíos legales:** quién puede acceder a estos datos, cómo se resguarda confidencialidad
- **Desafíos institucionales:** acuerdos institucionales para que la información pueda ser empleada con fines estadísticos, necesidad de registrar la información conforme a estándares, introducir identificador único (registros administrativos)
- **Desafíos de aseguramiento de calidad:** cómo se asegura la calidad de datos producidos por actores no tradicionales (más aún si son del sector civil o privado y no son entidades gubernamentales), cómo puede la ONE asegurar el cumplimiento de estándares
- **Desafíos técnicos:** algunas fuentes de datos requieren de conocimientos especializados para su procesamiento o tienen un elevado requerimiento computacional, que pueden no estar en las ONEs o en los SEN



**ALGUNOS TOPICOS A
DISCERNIR Y/O
REGLAMENTAR EN
TORNO AL USO
ADECUADO DE FUENTES
NO TRADICIONALES**

- **La legislacion sobre el uso de fuentes no tradicionales debe ser o no ligado a una legislacion nacional mas integrada?, por ejemplo en el caso de los registros.**
- **Como proteger los registros estadisticos claves que permiten los empalmes con otros registros?**
- **Que aspectos de gobernanza se deberian adaptar a cada “modelo de acuerdo interinstitucional”?, referente al uso de la fuente no tradicional.**
- **Un repositorio central y seguro para controlar el uso de datos administrativos podria ser un modelo a considerar? (como es el caso en Statistics Netherlands)**
- **Apalancar economias de escala deberia ser un objetivo clave al usar conjuntamente varias fuentes.**
- **Asuntos Claves: Seguridad y Soberania (Personal versus Nacional)**
- **Enfoque Integral PAPI: Proteger al individuo; Acceso abierto; derechos de Propiedad; Interoperabilidad**