

# Repensar la producción estadística oficial a través de la curaduría, integración y accesibilidad de datos

**Pedro Silva**  
**Mayo de 2024**

## Contexto

La producción de **estadísticas oficiales** depende principalmente de fuentes como **censos, encuestas por muestreo y registros**.

**Producir con calidad** requiere un conjunto amplio de:

conocimientos,

métodos, y

procesos

desarrollados en teoría y consolidados en la práctica de las organizaciones estadísticas.

## GSBPM → Modelo Genérico del Proceso Estadístico (MGPE)

	Etapas	Sub-procesos							
Gestión de la calidad / Metadatos / Paradatos	<b>Especificar requisitos</b>	Definir la información requerida	Consultar y confirmar la demanda	Establecer los resultados requeridos	Identificar conceptos	Comprobar la disponibilidad de datos	Preparar los términos de referencia		
	<b>Plan</b>	Especificar salidas ou resúmenes	Describir variables	Planificar el modo y la tecnología de recolección	Registro, marco y diseño de muestreo	Definir el método y el plan de crítica y imputación	Planificar sistemas y flujos de procesamiento		
	<b>Construir</b>	Desarrollar el instrumento de recolección	Configurar el flujo de trabajo	Elaborar componentes dos sistemas	Probar los sistemas de producción	Probar los sistemas estadísticos	Finalizar sistemas de producción		
	<b>Recolectar</b>	Seleccionar muestra	Preparar la recolección	Realizar la recolección	Finalizar la recolección				
	<b>Processar</b>	Integrar / organizar datos	Clasificar y codificar	Criticar y validar datos	Imputar datos faltantes / rechazados	Derivar variables y unidades de análisis	Calcular pesos de muestreo	Calcular estimaciones	Finalizar archivos de datos
	<b>Analizar</b>	Preparar borradores de outputs	Validar salidas	Analizar y explicar	Aplicar el control de confidencialidad	Finalizar salidas			
	<b>Diseminar</b>	Actualizar los sistemas de difusión	Producir productos de difusión	Gestionar la difusión de los productos de difusión	Promover productos de difusión	Gestionar la asistencia al usuario			
	<b>Archivo</b>	Definir reglas de archivo	Administrar repositorio	Conservar los datos y metadatos asociados	Descartar datos e metadatos de acuerdo con planos				
	<b>Evaluar</b>	Recopilar material para la evaluación	Realizar evaluación	Estabeecer plan de acción / mejora					

## Contexto

Las **estadísticas oficiales evolucionan lentamente**, ya que sus fuentes principales (censos y encuestas), requieren un largo desarrollo, consulta a los usuarios, experimentación e implementación para poder proporcionar datos que cumplan con los estrictos estándares de medición y calidad aplicables.

Las **necesidades de datos de la sociedad**, por otro lado, **evolucionan cada vez más rápidamente**, impulsadas por la innovación, la tecnología y la sed de datos para guiar las decisiones: **la "datificación" de la sociedad**.

# El gran reto

## **Special Section: Official Statistics**

---

**The Official Statistics Olympic Challenge:  
Wider, Deeper, Quicker, Better, Cheaper**

D. Tim HOLT

The American Statistician, vol. 61, 2007, febrero.

## Respostas de los INEs

Mejoras en la recogida de datos, a través de métodos más modernos y eficientes.

Aumento de muestras y rediseño de encuestas por muestreo.

Ampliación del uso de los registros administrativos.

Uso de modelos estadísticos, en particular, para producir estimaciones para dominios con muestras pequeñas.

Sin embargo, casi todas las iniciativas anteriores siguen centrándose en **fuentes de datos específicas**.

## Perspectivas

Para seguir siendo relevantes y atender las necesidades de los usuarios, las **estadísticas oficiales deben replantearse sus procesos de producción** estadística, pasando de los insumos y productos de una sola fuente (incluso los censos) a productos de datos:

- **de múltiples fuentes,**
- **de unidades múltiples,**
- **cubriendo temas y usos múltiples, y**
- **accesibles** a la gama más amplia posible de usuarios.

## Cómo avanzar

El rediseño de la infraestructura estadística del futuro va a requerir innovación y inversión en

- Selección,
- Curaduría,
- Integración y
- Infraestructura de acceso

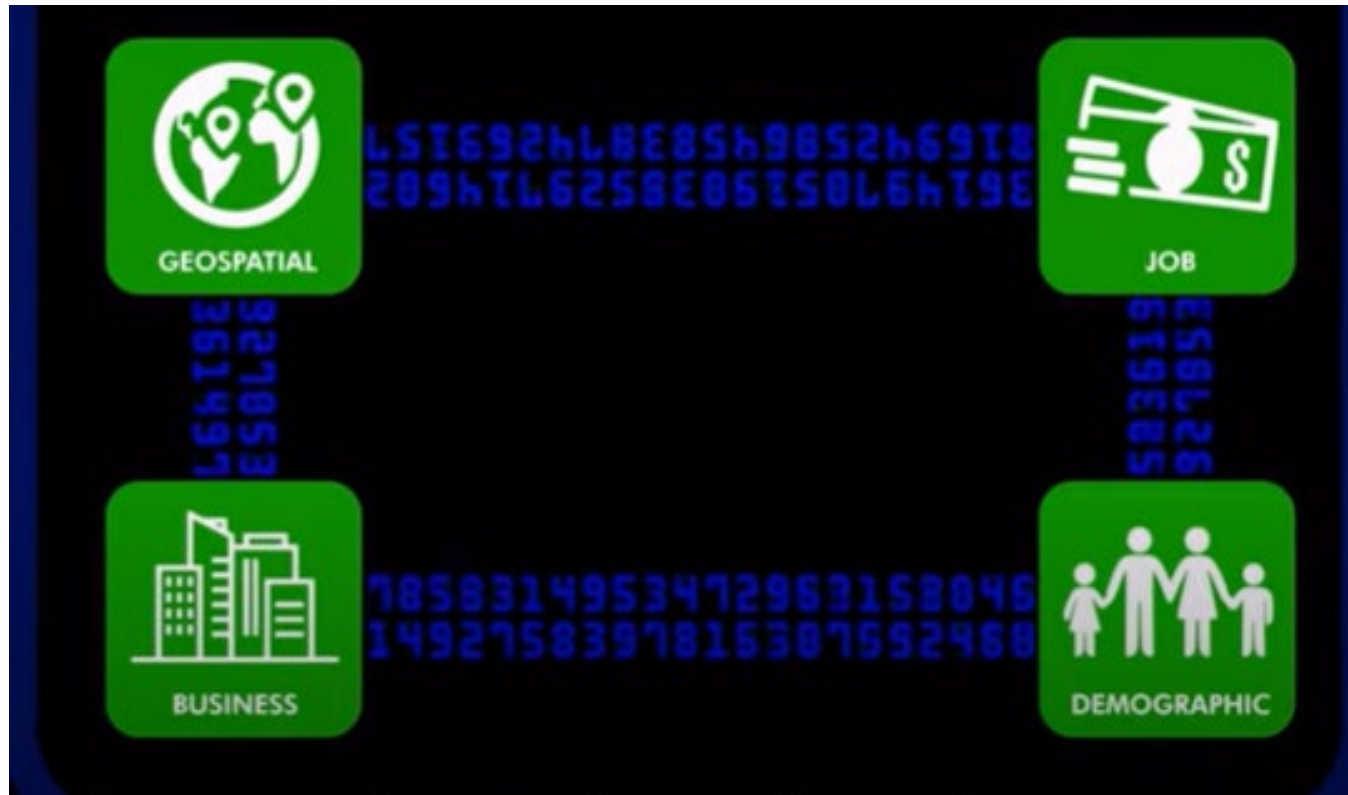
a los datos necesarios para responder a las demandas.

También será necesario innovar con los métodos, estándares, procesos y herramientas de apoyo a la producción estadística.



# Iniciativas inovadoras

## US Census Bureau “Frames Program”



## US Census Bureau “Frames Program”

La idea central es evolucionar a un modelo en el que exista un conjunto de registros integrados.

Cada registro se podrá vincular a los otros mediante claves de identificación únicas.

Y todos ellos verificados y complementados con la consulta de datos de los otros y de fuentes externas.

# US Census Bureau



# A 21<sup>ST</sup> CENTURY CENSUS CURATED DATA ENTERPRISE

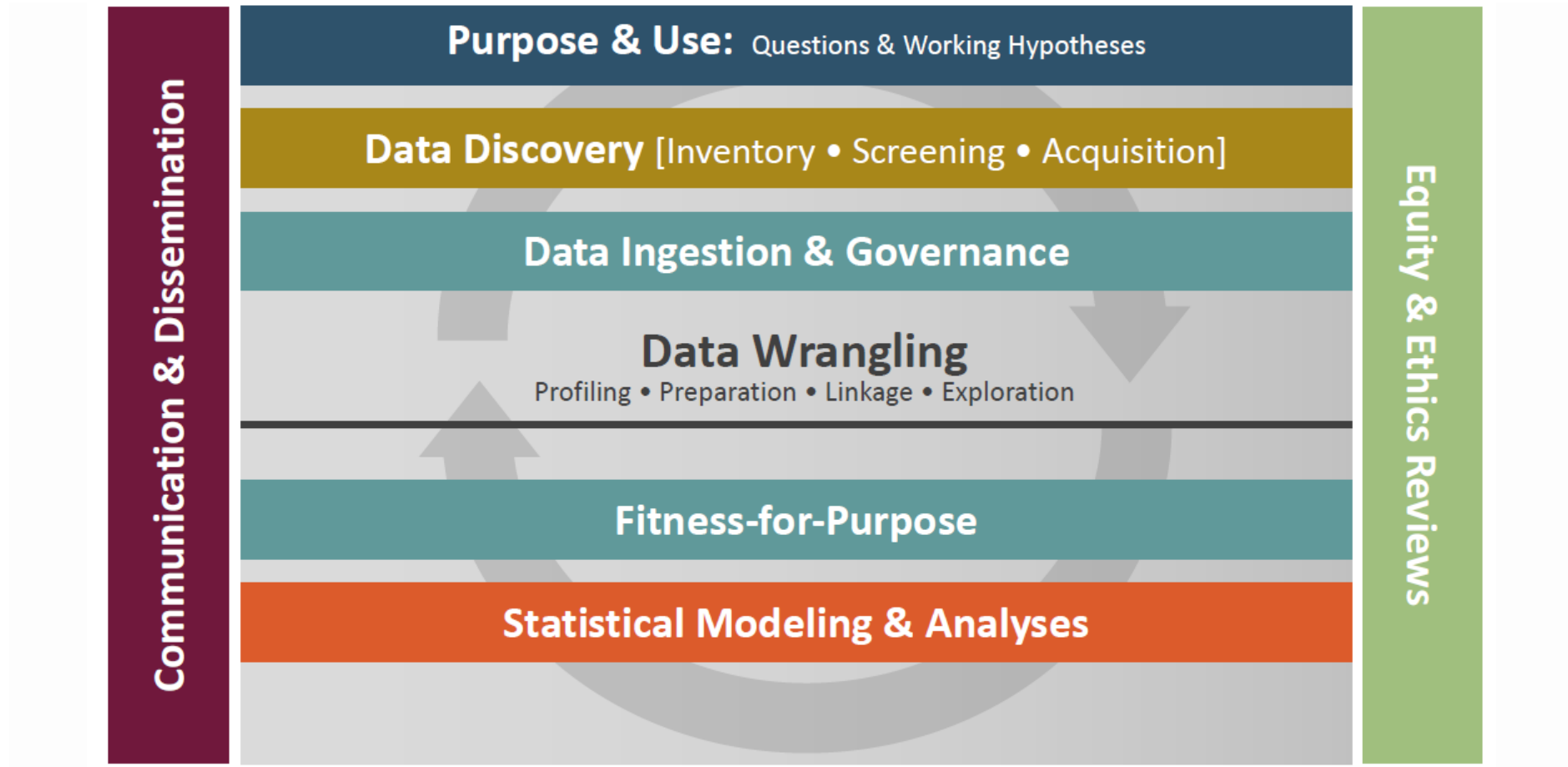
*A Bold New Approach to Create Official Statistics*

---

**Spring 2022 Report**

Sallie Keller, Ken Prewitt, John Thompson,  
Steve Jost, Chris Barrett, Sarah Nusser,  
Joseph Salvo, and Stephanie Shipp

# US Census Bureau



# Colaboração e compartilhamento de dados

## **Data collaboratives: possibilitando uma economia de dados saudável por meio de parcerias<sup>1</sup>**

Por Stefaan G. Verhulst<sup>2</sup>

Introdução

cobriria 150.000 vezes a distância da Amazônia. Se um *gigabyte* fosse um tijolo, um *zettabyte* seria equivalente a 258 Grandes Muralhas da China (3.873.000.000 de tijolos) (Barnett, 2016).

A dataficação de praticamente todos os aspectos de nossas vidas privadas e públicas apresenta oportunidades e desafios. Entre os desafios mais importantes, está o problema emergente das assimetrias de dados — a realidade incômoda apresentada pela escassez em meio a uma época de abundância sem precedentes. Embora nossa sociedade esteja inundada de dados, é cada vez mais claro que os dados e seus benefícios não são

## Data collaboratives

<https://cetic.br/pt/publicacao/ano-xv-n-4-colaboracao-e-compartilhamento-de-dados/>

## ***'Data collaboratives'***

El término se refiere a un modelo emergente de colaboración en el que los participantes de diferentes segmentos (agencias gubernamentales, empresas privadas, instituciones de investigación) intercambian datos para ayudar a resolver problemas públicos o de interés común.

## Modelos de colaboración

Interfaces públicas (p.ej. acceso abierto vía APIs)

Intermediario de confianza

*Pools* de datos

Parcerías para investigación y análisis

Premios y concursos

Generación de inteligencia

## Estadísticas Oficiais y Colaboración

No será posible avanzar sin colaboración.

Los INEs tendrán que acostumbrarse a colaborar más estrechamente con:

Academia;

Dueños de registros administrativos;

Empresas titulares de datos orgánicos;

Otros INEs;

Usuarios expertos.



**¡Gracias por su atención!**