



## MANUAL: TALLER DE METADATOS

---

Preparado por: Luz Mairym López Rodríguez, PhD  
Gerente de Oficina  
Instituto de Estadísticas de Puerto Rico  
Septiembre 2011



## TABLA DE CONTENIDO

I. Introducción a los metadatos .....	1
II. Tipos de metadatos .....	3
A. Metadatos de referencia .....	4
B. Metadatos estructurales .....	7
III. Herramientas de planificación para el desarrollo de metadatos .....	12
A. Modelo estadístico: Gerencia de calidad y metadatos .....	13
IV. Sistemas de información .....	16
V. Pensamientos finales .....	19
VI. Bibliografía .....	21
Anejo 1: Levels 1 and 2 of the Generic Statistical Business Process Model .....	23

## ÍNDICE DE TABLAS

Tabla 1: Fuentes de información en informes estadísticos .....	5
--	---

## ÍNDICE DE FIGURAS

Figura 1: Información de un informe estadístico en el Inventario de Estadísticas de Puerto Rico .....	4
Figura 2: Guía de Estadísticas Internacionales .....	7
Figura 3: Definición Delito Tipo I - Policía de Puerto Rico .....	8
Figura 4: Ejemplo índice de variables del Censo 2000 .....	9
Figura 5: Formulario para el Inventario de Estadísticas de Puerto Rico (Formulario) .....	10



## **I. Introducción a los metadatos**

---



Los metadatos “representan una serie de instrucciones o documentos que describen el contenido, contexto, calidad, estructura y accesibilidad de datos” (Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. & Stafford, S.G., 1997). Son los datos sobre los datos; las definiciones, descripción de procesos, parámetros y resultados operacionales que resumen operaciones estadísticas (Colledge, 1999). También se consideran como cada pieza de información necesaria para entender el significado del dato, teniendo presente armonizar las necesidades de los productores de los datos con los usuarios (Papageorgiou, H., Vardaki, M. & Pentaris, F., 2000).

El dato no se limita a un número o una palabra, sino a objetos concretos. Por ejemplo, imaginemos un conjunto de piezas para armar una mesa sin las instrucciones de cómo formarla, ¿podríamos armarla fácil y correctamente? Muy probablemente la mesa quedaría coja o con posibilidades de que no sea 100% servible. Otro buen ejemplo lo son las etiquetas en las latas de comida. Si las latas no tuvieran etiquetas sería casi imposible saber el contenido de las mismas y mucho menos conocer su valor nutricional. El menú del día sería toda una aventura cada vez que se abra una lata para cocinar.

Los metadatos explican los datos que se ven claramente, pero no necesariamente se tiene total conocimiento de los mismos. Es por eso que los metadatos son una herramienta, tanto para quien los prepara, como para el usuario. En el mundo de las estadísticas, los datos pueden representar realidades positivas o negativas, según el ángulo que se utilice para evaluarlos. Tener la mayor información de la procedencia de estos datos provee una perspectiva más amplia para determinar su significado. Publicar un dato sin la debida explicación de sus variables, margen de error, exclusiones, y metodología, entre tantas otras referencias, podría generar una controversia evitable. Los metadatos explican la estadística a los usuarios que no conocen la misma.

Aunque de primera instancia para algunas personas los metadatos no parezcan importantes, se debe hacer un esfuerzo e invertir los recursos necesarios. Los metadatos deben ser estables a través del tiempo y los sistemas de informática no deberían alterarlos. Afortunadamente, en la mayoría de los casos, la informática le da un valor añadido al estructurarlos. Probablemente las agencias gubernamentales y los municipios cuentan con metadatos aunque muchos no hayan sido identificados como tal. Al final de la lectura de este Manual podrán evaluar la magnitud de metadatos con que cuentan en sus lugares de trabajo.

Este Manual se divide en cuatro secciones. *Tipo de metadatos*, se enfoca en los metadatos de referencia y los estructurales. *Herramientas de planificación para el desarrollo de metadatos*, presenta algunas estrategias para desarrollar metadatos en proyectos. *Sistemas de información*, incluye el rol de los sistemas de información en el desarrollo de metadatos del Instituto. Finalmente cerramos con *Pensamientos Finales* sobre la información incluida en este Manual.



## **II. Tipos de metadatos**

---

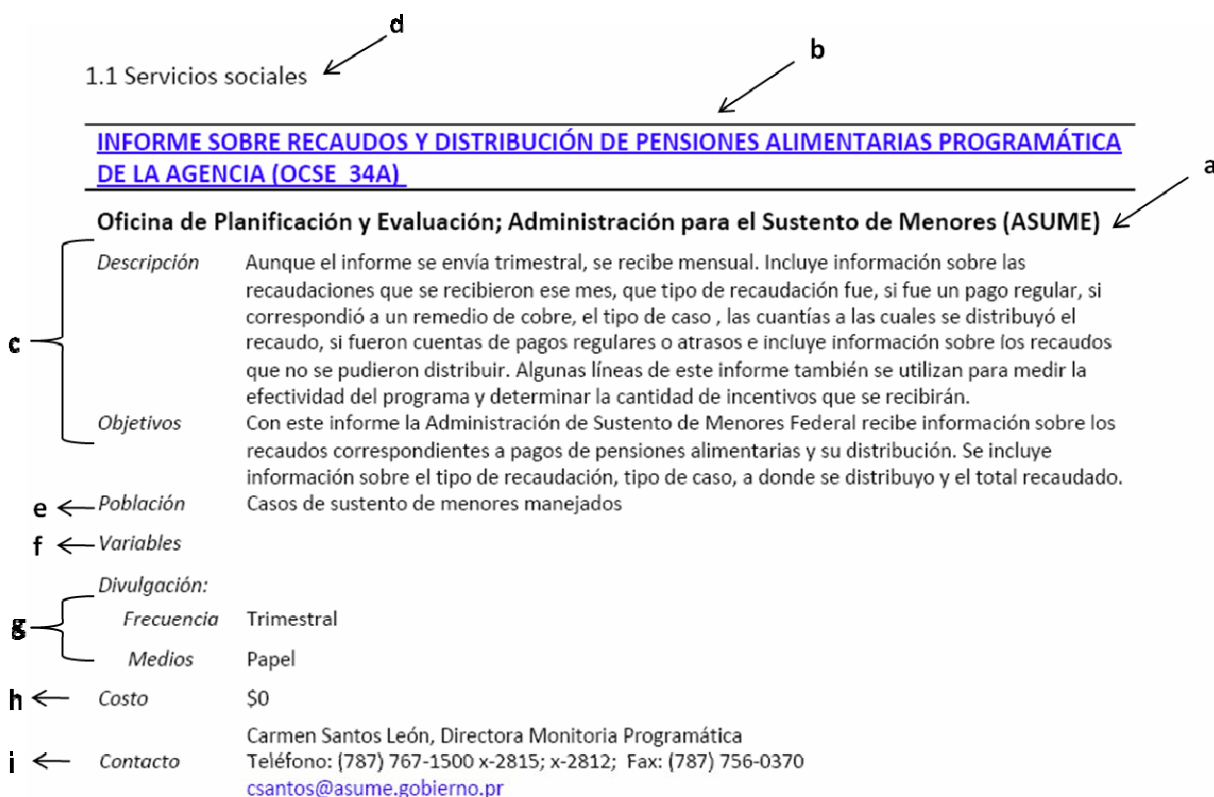


## A. Metadatos de referencia

Los metadatos de referencia son descriptivos asociados a cada operación estadística. Describe los contenidos, la metodología, la calidad de los datos, entre otros. Algunos ejemplos de este tipo de metadatos los veremos a continuación.

1. Inventario de Estadísticas: El Inventario de Estadísticas de Puerto Rico (Inventario) describe las actividades estadísticas de las agencias y municipios de Puerto Rico. Contiene varios elementos de referencia que podemos identificar en la Figura 1:
  - a. Agencias y municipios con informes estadísticos.
  - b. Informes estadísticos
  - c. Descripción y objetivos de los informes
  - d. Clasificación temática
  - e. Población
  - f. Variables
  - g. Divulgación
  - h. Costo
  - i. Contacto

**Figura 1:** Información de un informe estadístico en el Inventario de Estadísticas de Puerto Rico





2. Metodologías y fuentes de información: Esta referencia es necesaria para entender los procesos que se utilizan para generar una estadística. Uno de los criterios para pre-acreditar los informes estadísticos incluidos en el Inventario es “fuentes de información en informes estadísticos”. Esta información complementa la estadística. La Tabla 1 presenta ejemplos de fuentes de información de algunos informes estadísticos en el Inventario.

**Tabla 1:** Fuentes de información en informes estadísticos

---

**Informe Estadístico Anual del Negociado de Investigaciones Especiales**  
Departamento de Justicia

---

*Las estadísticas presentadas en este informe provienen de formularios o plantillas en formato Excel que se van recopilando mensualmente en las distintas divisiones del Negociado de Investigaciones Especiales, y se consolidan para análisis en la División de Planificación y Estadísticas. Las dependencias del NIE son: División Contra el Crimen Organizado; División de Integridad Pública; División de Protección; División de Asistencia a Víctimas y Testigos; División de Interpol; División de Coordinación, Enlace y Seguridad; y División de Capacitación y Apoyo Técnico.*

---

**Informe de Ingresos y Egresos en las Instituciones Correccionales (Mensual/ Anual)**  
Departamento de Corrección y Rehabilitación

---

*Las estadísticas presentadas en este informe provienen del registro administrativo Informe de Movimiento Diario realizado en todas las instituciones correccionales. Cada institución realiza el movimiento diario y lo envía a la Oficina de Desarrollo Programático indicando el total de la población ingresada y egresada cada día. Las variables principales del mismo son: ingresos por: (captura, sumariado, sentenciado, violador libertad bajo palabra y libertad a prueba, tránsito, traslado y reingreso). Las razones de egreso son: (multa, libertad bajo palabra, libertad a prueba, traslado, tránsito, fianza, cumplido, fuga, orden del tribunal, muerte y otros). Además incluye: promedio, máximo y mínimo para el informe anual.*

---

**Índice de Precios al Consumidor**  
Departamento del Trabajo y Recursos Humanos

---

Metodología IPC: La fórmula estadística que se utiliza para preparar el Índice de Precios al Consumidor es la del método agregativo de ponderación fija. Ésta, básicamente consiste en comparar, mensualmente, el valor que tienen las 503 categorías que componen la muestra (canasta) a los precios de cada mes, con el valor que tenían a los precios en diciembre de 2006 (periodo base), y multiplicar esa razón por cien para convertirla en el índice. Para más detalles vea en nuestro portal [www.estadisticaslaboralespr.com](http://www.estadisticaslaboralespr.com). Nota Técnica.



3. Reglamentos: La reglamentación es uno de los recursos utilizados para lograr la estandarización de las estadísticas. El Reglamento de los Criterios de Calidad de las Estadísticas, emitido por el Instituto, identifica los requisitos necesarios para que los sistemas de recopilación de datos y estadísticas en las agencias gubernamentales sean considerados de calidad. A su vez, la Carta Normativa Núm. 2011-01, también emitida por el Instituto, establece los requisitos mínimos para que informes estadísticos puedan ser pre-acreditados por el Instituto. Estos dos ejemplos, establecen aspectos necesarios que debe tener todo informe estadístico generado dentro de las agencias gubernamentales y municipios en Puerto Rico. Dentro de las agencias, se puede reglamentar muchos asuntos referentes a los informes estadísticos, tales como el contenido del informe, proceso y frecuencia de recopilación, publicación de los informes, formato, entre otros.
4. Procesos de encuestas y censos: Aunque los procesos de las encuestas y censos pueden considerarse parte de lo que es la metodología en el desarrollo de las estadísticas, la particularidad de los mismos hacen que se destaquen individualmente. Localmente, varias agencias realizan encuestas para generar estadísticas sobre diferentes áreas temáticas. Puede que internamente tengan los procesos utilizados para realizar la encuesta, pero el acceso al usuario para saber los mismos es limitado.

El *U.S. Census Bureau* (Negociado) es uno de los mejores ejemplos para validar la utilidad de los procesos en las encuestas, tanto para su personal, como para los participantes. En su

Ejemplos de encuestas:

- Consejo General de Educación: Informe de datos estadísticos de las instituciones educativas privadas no universitarias
- Departamento del Trabajo y Recursos Humanos - Encuesta del grupo trabajador; Encuesta de Índice de Precios; Encuesta de estadísticas de salarios
- Instituto de Estadísticas: Puerto Rico Manufacturing - Purchasing Managers' Index

portal de Internet se encuentra la sección *Are you in a survey?* con información sobre la veracidad de las encuestas, la información que el personal del Negociado debe proveerle al encuestado para legitimarlo, la política de privacidad de la información, la información de contacto del Negociado, entre otras (*U.S. Census Bureau*, n.d.). Para los que no tienen acceso a la Internet, el Negociado realiza campañas informativas, en especial en época del censo decenal. Es importante documentar estos procesos para garantizar la continuidad de los mismos, sin importar los cambios gerenciales y del personal a cargo, y también para revisar su validez al paso de los años.





## B. Metadatos estructurales

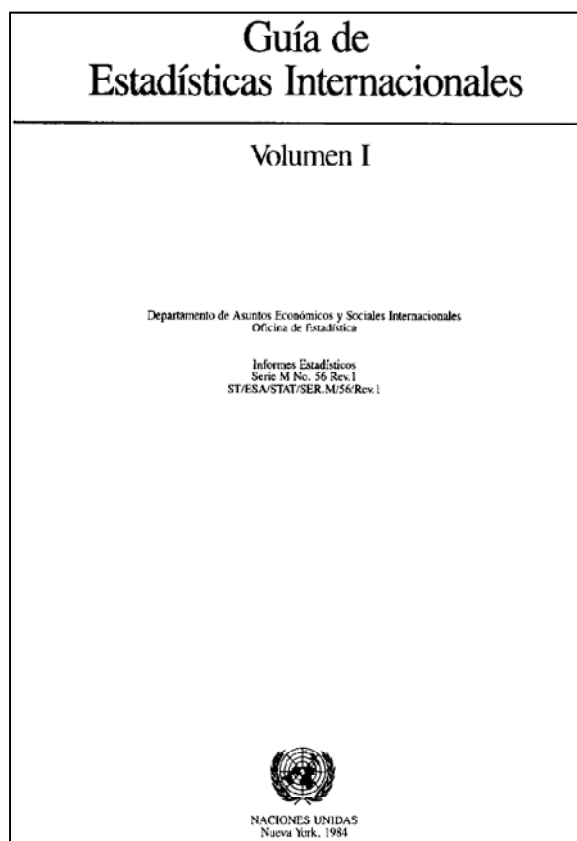
Los metadatos estructurales se utilizan para identificar, describir formalmente o recuperar los datos estadísticos, tales como variables, conceptos, conjunto de datos de descripciones técnicas, lista de códigos, formato de datos, rangos de valores, clasificaciones, etc. Estos son de utilidad al realizar el proceso estadístico desde su creación hasta su difusión. Los estándares, un tipo de metadato estructural, armonizan las unidades estadísticas, clasificaciones, variables, definiciones, contenido de instrumentos de recolección y otra terminología estadística (Colledge, 1999). Veamos ejemplos de metadatos estructurales.

1. Clasificaciones: Para realizar el Inventario de Estadísticas de Puerto Rico se utilizó la Guía de Estadísticas Internacionales - Volumen I, publicado por las Naciones Unidas (ver Figura 2). Esta Guía estandariza las clasificaciones de informes estadísticos por áreas temáticas de uso internacional.

De igual forma, podemos encontrar clasificaciones como el SIC y NAICS (clasificaciones de industrias), M49 e ISO (códigos de países), SOC (clasificación de ocupaciones), ICD (clasificación de enfermedades), entre otras.

2. Glosario de conceptos: Los glosarios permiten entender los datos estadísticos. El Informe estadístico “Delitos Tipo I” de la Policía de Puerto Rico (ver Figura 3), contiene una sección de definiciones incluyendo los siete delitos clasificados como Tipo I.

**Figura 2:** Guía de Estadísticas Internacionales





**Figura 3:** Definición Delito Tipo I - Policía de Puerto Rico

DEFINICIÓN DELITO TIPO I	
1. Asesinato y homicidio voluntario	<b>Definición-</b> La muerte intencional (no negligente) de un ser humano por otro.
2. Violación a la fuerza	<b>Definición-</b> La cópula con una persona, a la fuerza y/o contra la voluntad de esa persona; o no a la fuerza o contra la voluntad de la persona cuando la víctima es incapaz de dar su consentimiento debido a su incapacidad mental o física temporal o permanente (o debido a su juventud.)
3. Robo	<b>Definición-</b> Llevarse o intentar llevarse cualquier cosa de valor, en circunstancias de confrontación, que se encuentra bajo el control, custodia o cuidado de otra persona mediante la fuerza o amenaza de fuerza o violencia y/o poner a la víctima en temor de daño inmediato.
4. Agresión Grave	<b>Definición –</b> Un ataque ilícito de una persona contra otra en la cual el agresor utiliza un arma o la exhibe de manera amenazadora, o la víctima sufre una lesión corporal severa o grave, incluyendo huesos fracturados, pérdida de dientes, posibles lesiones internas, desgarraduras severas o pérdida de conocimiento aparentes.
5. Escalamiento/Allanamiento de Morada	<b>Definición –</b> La entrada ilícita a un edificio u otra estructura con la intención de cometer un delito mayor o un robo.
6. Apropiación Ilegal	<b>Definición-</b> Adquirir, transportar, llevar o montarse en y llevarse bienes que le pertenecen real o legalmente a otra persona.
7. Hurto de auto	<b>Definición-</b> El hurto de un vehículo de motor.  Se deben clasificar como hurto de auto de vehículos de motor todos los casos en que las personas que personas que no tienen acceso legal a automóviles se los llevan y posteriormente los abandonan. Se debe incluir el llevarse un vehículo sin permiso del dueño para realizar paseos alocados.

Los glosarios de conceptos son de utilidad para organismos grandes con varias unidades o subsidiarias, ya que se estandarizan y se mantienen uniformes los conceptos. Estos conceptos se utilizan en encuestas, formularios para recopilar datos, entre otros. Como ejemplo, la Universidad de Puerto Rico desarrolló en el 1998 un Glosario de Términos de los términos estadísticos que utilizan todas sus unidades con el “fin de facilitar la producción de informes estadísticos mecanizados” (Universidad de Puerto Rico, 1998, p.6).

- Variables: De acuerdo al Glosario de clasificación de términos de las Naciones Unidas (n.d.), una variable es una de las características que se observa en una unidad y que puede estar definida como un conjunto de valores. Las encuestas, los informes estadísticos y cualquier instrumento de recopilación de datos se componen de diversas variables cuantitativas y cualitativas. Estas variables deben estar definidas y codificadas para recopilar bien los datos y posteriormente analizarlos y que el usuario los entienda una vez sean publicados.



La Figura 4 presenta parte del diccionario de variables publicado por el Negociado en el informe de documentación técnica, para beneficio de los usuarios que acceden los archivos de datos y la documentación técnica (ver figura 4). Muy probablemente, sin esta referencia no se podría entender las codificaciones de las variables utilizadas en los archivos del Negociado.

**Figura 4:** Ejemplo índice de variables del Censo 2000

<b>INDEXES</b>		
Alphabetical Index by Variable Name (Housing Unit Record)		
Variable name	Character location	Description
ACRES	H138	Acreage
ACRESA	H139	Acreage Allocation Flag
AGSALES	H140	Sales of Agricultural Products in 1999
AGSALESA	H141	Sales of Agricultural Products in 1999 Allocation Flag
AREATYP1	H42-43	Metropolitan Area: Super-PUMA Relationship to MA
BEDRMS	H124	Number of Bedrooms
BEDRMSA	H125	Number of Bedrooms Allocation Flag
BLDGSZ	H115-H116	Size of Building
BLDGSZA	H117	Size of Building Allocation Flag
BUSINES	H136	Commercial Business on Property
BUSINESA	H137	Commercial Business on Property Allocation Flag
CKITCH	H128	Complete Kitchen Facilities
CKITCHA	H129	Complete Kitchen Facilities Allocation Flag

Una forma factible de definir las variables es al momento de formación del instrumento de recopilación de datos, ya que es cuando se determinan las necesidades, el por qué y para qué de cada pregunta. Veamos en la Figura 5 una porción del Formulario para el Inventario de Estadísticas de Puerto Rico.



**Figura 5:** Formulario para el Inventario de Estadísticas de Puerto Rico (Formulario)

**Frecuencia de divulgación del informe**

☐ Diaria  
La frecuencia de divulgación se refiere a la cantidad de veces que se da a conocer el informe estadístico. Si la publicación del informe no ha permitido establecer una regularidad, puede marcar "Otro" e indicar "variable" seguido por las últimas dos fechas de divulgación del informe. Por ejemplo: variable, julio 2005 y enero 2007.

☐ Semanal  
☐ Bimensual  
☐ Trimestral  
☐ Anual  
☐ Otro:

**9) Período de referencia de datos**

☐ Día  
☐ Semana  
☐ Bisemana  
☐ Quincena  
☐ Mes  
☐ Trimestre  
☐ Otro:

**Año:**  
☐ natural  
☐ escolar  
☐ fiscal local (jul. – jun.)  
☐ fiscal federal (oct. – sep.)

**10) Solicitud de informe**

☐ En persona  
☐ Correo postal  
☐ Teléfono  
☐ Fax  
☐ Correo electrónico

**11) Obtención de informe**

☐ En persona  
☐ Correo postal  
☐ Fax  
☐ Correo electrónico  
☐ En portal

**12) Medios de divulgación de informe**

☐ En papel  
En archivo electrónico:  
☐ MS Excel  
☐ Texto  
☐ PDF (readable)  
☐ PDF (scan)

**13) Costo del informe**

☐ No  
☐ Sí, ¿cuánto?

**14) Métodos de pago**

☐ Efectivo  
☐ Sello de colecturía  
☐ ATH  
☐ No aplica

☐ Tarjeta de crédito  
☐ Giro  
☐ Cheque

Las variables en esta porción del Formulario son: frecuencia de divulgación del informe, periodo de referencia, solicitud de informe, obtención de informe, medios de divulgación de informe, costo del informe y métodos de pago. Al utilizar el formulario en el programa Adobe, las definiciones de algunas de las variables se definen en un recuadro amarillo al colocar el cursor sobre el título. Este metadato ayuda al encuestado a entender lo que se pretende con la pregunta y a su vez minimizar los errores en las respuestas.

Otras variables en el mismo Formulario no se definen, por lo que deberían incluirse las mismas. Los diseñadores del instrumento de recopilación necesitan definir las variables en conformidad con el objetivo de la encuesta. Los encuestados tienen que entender las variables (preguntas) para responder lo mas preciso posible. Los usuarios externos necesitan conocer las definiciones para la aplicación apropiada al utilizar datos para sus proyectos e investigaciones.

- Operación estadística: Todas las estadísticas se generan utilizando un proceso matemático. Estos pueden ser sencillos como complejos, por lo que es vital tener la referencia del cálculo



realizado. Un ejemplo es la fórmula matemática del estimador de ventas utilizada por la Compañía de Comercio y Exportación para generar el Informe de Ventas al Detalle (InfoVentas). La Sección 2.3 del InfoVentas de diciembre de 2005, explica al detalle esta fórmula y la metodología utilizada. Esta herramienta es de utilidad al momento de replicar la fórmula, cambiar la persona a cargo de elaborar el cálculo o sencillamente analizar la estadística generada.

También podemos encontrar en el Informe Anual de Estadísticas Vitales 2006 (Departamento de Salud, 2010) las fórmulas para calcular las tasas estadísticas de defunción, nacimientos, matrimonios y divorcios. Para sacarle un mayor provecho a estas fórmulas, se podrían complementar con las definiciones de cada variable utilizada.



### **III. Herramientas de planificación para el desarrollo de metadatos**



El *Código de buenas prácticas de las estadísticas europeas* (2005) contiene cuatro áreas pertinentes en el proceso de desarrollo de metadatos: metodología sólida, procedimientos estadísticos adecuados, coherencia y comparabilidad y accesibilidad y claridad. Crear metadatos no debe ser un proceso para cumplir con un cometido, sino para lograr que las buenas prácticas identificadas se cumplan.

1. Metodología sólida: Tener procedimientos para garantizar que se aplican coherentemente conceptos, definiciones y clasificaciones estándares en toda la estadística.
2. Procedimientos estadísticos adecuados: Las definiciones y los conceptos utilizados para efectos administrativos deben aproximarse bastante a los seguidos para efectos estadísticos. Probar los instrumentos de recopilación de datos.
3. Coherencia y comparabilidad: Las estadísticas se recopilan sobre la base de normas comunes respecto al alcance, las definiciones, unidades y clasificaciones en las distintas encuestas y fuentes.
4. Accesibilidad y claridad: Los metadatos están documentados con arreglo a sistemas de metadatos.

Los metadatos integran tres áreas importantes: datos estadísticos (conceptos, variables, clasificaciones, etc.), proceso estadístico (producción, recogida, almacenaje, análisis, difusión, etc.) y herramientas fáciles y aptas para la automatización del proceso (sistemas de procesamiento de datos, formularios, cuestionarios, etc.). Estos deben coexistir para facilitar el desarrollo de modelos estándares que fomenten el intercambio y reutilización de los metadatos.

Dado que los metadatos deben promover la accesibilidad a los usuarios y facilitar la producción de las estadísticas, desarrollarlos requiere de un proceso de planificación para lograr una gerencia de calidad.

#### A. Modelo estadístico: Gerencia de calidad y metadatos

El Modelo genérico del proceso de negocio estadístico, aprobado como referencia en marzo de 2009, es el producto del trabajo realizado por la Comisión Económica para Europa de la Organización de las Naciones Unidas, Eurostat y la Organización para el Desarrollo y Cooperación Económica (ver Anejo 1). Este Modelo, integra la gerencia de calidad y la gerencia de metadatos.



1. Gerencia de calidad: Es el proceso de planificar, implementar, evaluar y mejorar.
2. Gerencia de metadatos: Es la integración del manejo del metadato, la autoridad del metadato, relación con los ciclos estadísticos y los procesos, y los usuarios.

Este Modelo establece un proceso de nueve fases similares al desarrollo de gerencia de proyectos. Estas fases son flexibles y no todos los niveles aplican a todos los casos. Esto significa que el modelo puede ser lineal o variable. A continuación veremos cada una de estas fases.

1. Especificación de necesidades: En esta etapa se determinan las necesidades a través de consultas, se establecen los objetivos de producción, se identifican los conceptos, se evalúa la disponibilidad de los datos y se prepara el caso.
2. Diseño del producto estadístico: Esta fase requiere diseñar el producto, la descripción de las variables, la metodología de recolección de datos, la muestra y el marco muestral, la metodología del proceso estadístico y del sistema de producción y el flujo de trabajo.
3. Desarrollo y proceso de producción del instrumento de recogida: En esta etapa se desarrolla el instrumento de recolección de datos y los componentes del proceso. El instrumento debe tener los controles necesarios, pero no demasiados, ya que a mayor número de controles, mayor la dificultad para recibir respuestas. Además, se continúa la configuración del flujo de trabajo, se prueba el sistema de producción y el proceso estadístico y se finaliza el sistema de producción. Esto requiere definir la base de datos para que los conceptos iguales tengan los mismos códigos. En casos en que las encuestas tengan mucha frecuencia, es probable que esta etapa se realice la primera vez que se desarrolle el instrumento y sólo requiera de revisiones periódicas.
4. Recogida de los datos: Para que haya una recogida debe seleccionarse una muestra, preparar el proceso de recogida y realizar la misma.
5. Procesamiento de la información: Esta fase puede ir en paralelo con el análisis. En la misma se integran, clasifican, codifican y depuran los datos. depuraciones puedes ser semiautomáticas, donde se hace una detección automática de la falta de respuesta o error y se hace una corrección manual, o automáticas, donde el proceso de detección y corrección es automático (González Collar & Rueda Clausell, n.d.). Finalmente, se imputan los datos que faltan o tienen errores, se hacen los cálculos de datos agregados y se revisa la base de datos para análisis.





6. Análisis: El análisis es una de las etapas más importantes, ya que se prepara el borrador de la información recibida, se identifican los datos relevantes para explicar e interpretar y se prepara lo que posteriormente se difundirá.
7. Difusión: En la difusión se promueve la información estadística generada en tablas o informes, a través de comunicados y notas de prensa, el portal de Internet, suscripciones, entre otros.
8. Archivo: Al finalizar con la elaboración de datos y metadatos, estos deben archivar en un repositorio y se recomienda reglamentarlo para establecer el contenido, catalogación, mantenimiento, entre otros.
9. Evaluación: Finalmente la evaluación sirve para mejorar, reutilizar y establecer un plan de acción que garantice la calidad de los datos y metadatos.



## IV. Sistemas de información

---



La informática es un componente técnico que provee un valor añadido a la estructura de los datos. Esta ayuda en la integración de diferentes bases de datos, difusión adecuada y herramientas de búsquedas. Sin embargo, en muchas ocasiones los productos son costosos y se continúan utilizando modelos obsoletos que no ayudan a agilizar el procesamiento de datos estadísticos.

Dado que mi especialidad no es en el área de sistemas de información, entendí pertinente entrevistar a Rodolfo Pagán, Gerente de Sistemas de Información (GSI) del Instituto, sobre el rol de los sistemas de información en el desarrollo de datos del Instituto. En esta entrevista se identificaron tres áreas principales: bases de datos, inventario de estadísticas y herramientas de encuestas.

1. En el portal de Instituto contamos con bases de datos clasificadas por áreas temáticas. Dentro de la base de economía encontramos a LAUS y CES que proveen datos de Puerto Rico, pero se alimentan de bases de datos federales. Esto requiere:
  - a. obtener los datos de las fuentes federales en formato de texto;
  - b. Si los datos no tienen referencia sobre la codificación que generó los mismos, convertir los datos en texto para crear unas claves y codificaciones y;
  - c. extraer los datos en tablas
  - d. Incluir los enlaces de diccionario de variables (metadato estructural)

Estas bases son parte del proyecto “Mega Base” que el Instituto está desarrollando y se encuentra en la fase inicial. Este proceso, aunque manual, permite mantener un control.

2. El Inventario de Estadísticas de Puerto Rico se desarrolló creando una forma en PDF con el programa Adobe Acrobat Professional (Adobe). Esta forma se envió a todas las agencias y municipios, recibiendo la mayoría digitalmente. Las que fueron entregadas en papel o escaneadas se transcribieron en el documento en PDF. Una vez cerrado el periodo de recibo de los formularios, se exportó la data a Excel, conformando una base de datos amplia con variables definidas. Esta base se revisó para identificar errores y corregirlos, y generar el primer Inventario de Estadísticas de Puerto Rico. De no tener la herramienta de creación de formas y exportación de datos que provee Adobe, probablemente el trabajo hubiese sido más intenso.

La integración de esa base en el portal en la Internet del Instituto requirió programar. En vista de que muchos productos de programación son costosos, se utilizaron recursos gratuitos de código abierto: Dot Net Nuke (gestor de contenido); Visual Basic y C Sharp (herramientas de programación) y SQL Express (manejador de bases de datos). Estas herramientas permiten



que la información del Inventario de Estadísticas esté disponible al usuario en el portal en el Internet del Instituto y puedan realizarse búsquedas por tema, agencia y palabra clave.

3. Para desarrollar la Encuesta de Manufactura en línea se utilizaron dos herramientas de código abierto: “Lime Survey” y MySQL. El GSI del Instituto, creó la interface de la encuesta utilizando Lime Survey. Las encuestas se envían al correo electrónico de las personas a encuestarse con una codificación única (“llave”) para que el sistema reconozca a los que responden. De ser necesario se realizan envíos de recordatorio sólo a los que no han respondido. Las respuestas de la encuesta van directamente a una base de datos de MySQL para el análisis e interpretación de los mismos.

Estos tres ejemplos requieren de metadatos de variables y codificaciones que faciliten la transferencia de datos para el usuario. A parte del programador, que probablemente sea parte del personal en nuestras agencias gubernamentales y municipios, las herramientas son gratuitas, minimizando el impacto presupuestario.

El proyecto de metadatos en el Instituto está en su etapa inicial, por lo que se estará desarrollando prospectivamente. Al diseñarlo es importante tomar en cuenta que los metadatos deben ser estables en el tiempo y los sistemas de informática no deben alterarlos, sino complementarlos.



## V. Pensamientos finales

---



A través de este taller identificamos lo que son los metadatos y enfocamos la discusión en dos tipos de estos. Identificamos herramientas de planificación para integrar los metadatos en el proceso de generación de datos. También exploramos el complemento de los sistemas de información para fortalecer los metadatos.

El conocimiento de los datos puede que esté incompleto si no se entiende el proceso utilizado para obtenerlo. Michener, et. al, (1997) presenta como una limitación el utilizar datos entre colegas o conocedores de un tema, sin tener las instrucciones que ayuden a usar e interpretar los mismos correctamente. Es por eso que los metadatos aportan grandemente en el mundo de las estadísticas. Algunos beneficios son:

1. **Preservación histórica:** La memoria es corta, por lo que los metadatos ayudan a subsanar el olvido con el transcurso del tiempo y la ausencia de quienes generaron los datos (Michener, et. al, 1997). Esta limitación la observamos en muchas facetas, pero en asuntos tan complejos como la generación de datos, se debería tener mayor control para que no se pierda el conocimiento que sólo la mente humana guarda.
2. **Estudios longitudinales:** En estudios longitudinales, en los cuales se evalúan las muestras por un periodo a largo plazo, son esenciales los metadatos. En especial la metodología, variables y definición de conceptos.
3. **Reúso de datos previamente recopilados:** En muchas ocasiones se utilizan datos del pasado (años, décadas) para evaluar cambios e identificar patrones de conducta. Antes de la era digital, muchos datos estaban en papel, pero hasta así se pueden utilizar si se tienen todas las referencias necesarias. Ahora que estamos en la era digital el proceso se aliviana, agilizando el reúso de los datos.
4. **Reducción de duplicidades:** La falta de normas y estandarización puede ser un factor para que se dupliquen procesos, que pueden ahorrar tiempo y dinero. Por tal razón, el organizar los metadatos, puede generar mejores estrategias para que no se tengan que solicitar los mismos datos mediante diferentes métodos.

El estudio del metadato es un área novel con implementación limitada en Puerto Rico, pero tiene el potencial de aportar, siempre que se utilice con un propósito y no por el simple hecho de tenerlos.

***¡Éxito!***



## VI. Bibliografía

---

- Colledge, M.J. (1999). Statistical Integration through Metadata Management. *International Statistical Review*. 67(1), pp. 79-98.
- Compañía de Comercio y Exportación. (2005). InfoVentas: El Informe de Ventas al Detalle de Puerto Rico. Documento #1.
- Departamento de Salud. (2010). Informe Anual de Estadísticas Vitales de Puerto Rico, 2006. Secretaría Auxiliar de Planificación y Desarrollo. San Juan, PR.
- Eurostat & Instituto Nacional de Estadística (2005). Código de buenas prácticas de las estadísticas europeas. Aprobado por el Comité del Programa Estadístico el 24 de febrero de 2005.
- González Collar, A. & Rueda Clausell, A. (n.d.) Recogida y tratamiento informático de datos. En *Fundamentos y prácticas de las encuestas a los hogares, 1ª edición*, Módulo 2.
- Instituto de Estadísticas de Puerto Rico. (2010). Formulario para el Inventario de Estadísticas de Puerto Rico. San Juan, PR.
- Instituto de Estadísticas de Puerto Rico. (2010). Inventario de Estadísticas de Puerto Rico. San Juan, PR.
- Instituto Nacional de Estadísticas. (2011). Modelo Genérico de Producción de Estadística desarrollado en el marco de Naciones Unidas. Seminario sobre estándares y metadatos, realizado del 23 al 27 de mayo de 2011 en el INE de España.
- Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. & Stafford, S.G. (1997). Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications*. 7( 1), pp. 330-342.
- Naciones Unidas (1984). Guía de Estadísticas Internacionales: Volumen I. Departamento de Asuntos Económicos y Sociales Internacionales. Nueva York, USA.
- Papageorgiou, H., Vardaki, M. & Pentaris, F. (2000). Data and Metadata Transformations. *Research in Official Statistics*. 3(2), pp.27.43.
- Policía de Puerto Rico (2011, Julio). Definición: Delitos Tipo I. División de Estadísticas de la Criminalidad. San Juan, PR.
- UNECE Secretariat. (2009). Generic Statistical Business Process Model. Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS).



Unites Nations. (n.d.). UN Glossary of Classification Terms. Descargado el 15 de agosto del portal en el Internet de la Organización de las Naciones Unidas, [http://unstats.un.org/unsd/class/family/glossary\\_short.asp#V](http://unstats.un.org/unsd/class/family/glossary_short.asp#V).

Universidad de Puerto Rico. (1998). Glosario de términos. División de Impresos UPR. San Juan, PR.

U.S. Census Bureau. (n.d.). Are You in a Survey? Descargado el 15 de agosto del portal en el Internet de US Census Bureau, [http://www.census.gov/survey\\_participants/household\\_surveys/](http://www.census.gov/survey_participants/household_surveys/).

U.S. Census Bureau. (2003). Public Use Metadata Sample: 2000 Census of Population and Housing. Technical Documentation. United States.

Vale, S. (2009). Generic Statistical Business Process Model. United Nations Economic Commission for Europe.





## Anejo 1: Levels 1 and 2 of the Generic Statistical Business Process Model

