

# Red Nacional de Metadatos.

## Herramienta para el uso informado de los datos

**Marco Antonio Gutiérrez Romero**

El uso cada vez más frecuente de herramientas poderosas para diseminar datos ofrece la posibilidad de transferir grandes volúmenes de ellos en tiempo real, lo cual propicia que un mayor número de personas haga uso de la información; lo anterior, puede representar un riesgo en la interpretación de los datos si éstos no van acompañados de una documentación o metadato completo y de calidad; por ello, es necesario proveer a los usuarios de una documentación más detallada y de fácil acceso. La Red Nacional de Metadatos (RNM) es la herramienta que el Instituto Nacional de Estadística y Geografía (INEGI) ha implementado como medio de divulgación y consulta de metadatos de los proyectos estadísticos generados por éste; constituye el resultado de un conjunto de esfuerzos realizados para contar con un repositorio sistematizado de la documentación de operaciones estadísticas de diversos tipos (censos, encuestas y registros administrativos, por ejemplo) que han sido documentadas mediante la adopción de estándares internacionales.

**Palabras clave:** metadatos, DDI, estándares, microdatos, documentación.

The increasingly frequent use of powerful tools to disseminate data offers the possibility of transferring large amounts in real time, a greater number of people can use information; this may pose a risk the interpretation when these don't go with a complete quality metadata. It's a necessity to provide data users with more detailed documentation of easy access. The Metadata National Network (RNM by its Spanish acronym) is the tool that INEGI has implemented as a means of dissemination and query for metadata of statistical projects generated by the institute. RNM is the result of a series of efforts to have a systematic repository of statistical operations of various types, such as censuses, surveys and administrative records that have been documented using the international standards.

**Key words:** metadata, micro data, documentation, standards.



Fishermen Under Nets at Sunset. Gajim/

## ¿Qué son los metadatos?

La definición etimológica nos dice que son "datos acerca de los datos"; en general, describen o dicen algo sobre otro objeto de información. Podríamos decir que un metadato es información sobre las características de los datos que permiten entender *lo que representan* para que puedan ser compartidos y explotados de manera eficaz por todo tipo de usuarios a lo largo del tiempo.

## ¿Por qué son importantes?

Las estadísticas siempre han sido una herramienta valiosa para sustentar la toma de decisiones en la búsqueda de crear mejores condiciones de vida para la sociedad; sin embargo, los datos sin un contexto por lo regular son subutilizados o se corre el riesgo de llegar a conclusiones no del todo preci-

sas; por ello, la documentación de éstos tiene gran importancia para todos los usuarios de los datos, además de ser la base para la creación de una cultura estadística.

Los metadatos proporcionan a los usuarios información sobre los datos existentes, describiendo los procesos de recolección, procesamiento y evaluación que se utilizaron en su generación, así como las formas de acceder a ellos, con el fin de que los usuarios puedan identificar, localizar y consultar los que atiendan de mejor manera sus necesidades. En otras palabras, los metadatos ayudan al usuario a transformar los datos estadísticos en información, de ahí la importancia de la adopción de una terminología común y un conjunto de elementos estándar documentados con apego a normas bien establecidas y de aceptación internacional. La figura 1 ejemplifica la importancia que éstos tienen para garantizar un uso informado de los datos estadísticos.

Figura 1

## Importancia de los metadatos para el uso informado de datos\*

**Existe un riesgo alto al consumir productos sin saber qué contienen**



**Riesgo de malinterpretar datos sin metadatos**

CONAGU	DRENAJE	COMBUST	ELIBAS
5	1	1	1
5	1	1	1
6	1	1	1
7	5	3	4
6	2	1	1
6	2	1	3
	5	3	7
	5	3	7
9	5	1	3
	3	3	1
5	2	1	3
5	2	1	3
5	2	1	1

**Consumo responsable de productos debidamente documentados**



**Uso informado de los datos con metadatos**

### Drenaje (DRENAJE)

Archivo: Vivienda\_Ampliado

Categorías

Valor Categoría

- 1 Red pública
- 2 Fosa séptica
- 3 Tubería que va a dar a una barranca o grieta
- 4 Tubería que va a dar a un río, lago o mar
- 5 No tiene drenaje
- 9 No especificado

Proceso de documentación de metadatos

\* La idea de *The Bean Data Analogy* fue tomada del documento *A Manager's Introduction to Adobe extensible Metadata Platform, The Adobe XML Metadata Framework*, disponible en: [www.adobe.com/products/xmp/pdfs/whitepaper.pdf](http://www.adobe.com/products/xmp/pdfs/whitepaper.pdf)

## Estándares internacionales de metadatos

En años recientes, tanto a nivel nacional como internacional se ha enfatizado sobre la necesidad de contar con información documentada en forma armonizada, con apego a normas establecidas y de aceptación a nivel mundial.

Son lineamientos para mejorar la comparabilidad entre las naciones, pudiendo ser usados para armonizar los procesos de generación, compilación y diseminación de información estadística. Son desarrollados por acuerdo de expertos técnicos de diversos países bajo cooperación internacional y se constituyen en recomendaciones internacionales por su carácter no vinculante.

Para facilitar los procesos de armonización de las estadísticas oficiales, se ha propuesto una gran cantidad de iniciativas (ver tabla 1) sobre los metadatos estadísticos.

## Herramientas para la gestión de microdatos

Como una respuesta a la necesidad de los usuarios de contar con información documentada y con la finalidad de atender los objetivos del Programa Acelerado de Datos<sup>1</sup> de la Red Internacional de Encuestas de Hogares (IHSN, por sus siglas en inglés),<sup>2</sup> el Grupo de Datos del Banco Mundial ha elaborado el conjunto de Herramientas para la Gestión de Microdatos (HGM), las cuales están basadas en tecnología Nesstar y fueron desarrolladas por el Norwegian Social Science Data Services<sup>3</sup> (NSD) y tienen como propósito promover la adopción de los estándares internacionales de la Iniciativa

1 [www.ihsn.org/adp/](http://www.ihsn.org/adp/)

2 [www.surveynetwork.org/home/index.php?](http://www.surveynetwork.org/home/index.php?)

3 [www.nsd.uib.no/nsd/english/index.html](http://www.nsd.uib.no/nsd/english/index.html)

de Documentación de Datos (DDI, por sus siglas en inglés) y la Dublin Core Metadata Initiative (DCMI) para la documentación, difusión y preservación de datos, así como fomentar las mejores prácticas entre los productores de datos de los países en vías de desarrollo.

## Sistemas de metadatos en el INEGI

En el Instituto, la necesidad de documentación fue identificada hace varios años; por ello, se desarrollaron los siguientes proyectos de documentación:

### Registro Estadístico Nacional (REN)

Es un acervo de información estadística de cobertura nacional cuyo objetivo es integrar los metadatos que permitan identificar, caracterizar y clasificar a las unidades administrativas con funciones estadísticas del sector público federal y estatal, así como los proyectos estadísticos que atienden y los productos resultantes.

Tabla 1

### Algunas iniciativas internacionales relacionadas con metadatos estadísticos

Sistema	Breve descripción
Statistical Metadata (METIS) del Grupo de Trabajo sobre Metadatos Estadísticos de la UNECE	Mantiene una base dinámica de conocimiento para facilitar la armonización de estructuras y modelos de datos para metadatos estadísticos en el contexto del procesamiento y la difusión de la información estadística. <a href="http://www.unece.org/stats/archive/04.01d.e.html">http://www.unece.org/stats/archive/04.01d.e.html</a>
Standard Data and Metadata Exchange (SDMX)	Desarrollo de estándares técnicos para el intercambio de datos y metadatos, con énfasis en datos agregados. <a href="http://sdmx.org/">http://sdmx.org/</a>
Data Documentation Initiative (DDI)	Es un esfuerzo por establecer un estándar internacional basado en XML para el contenido, presentación, transporte y preservación de datos. <a href="http://www.ddialliance.org/">http://www.ddialliance.org/</a>
Dublin Core Metadata Initiative (DCMI)	Desarrolla estándares de metadatos para datos de un amplio marco de tipos de datos y modelos de negocios. Está integrado por bibliotecas nacionales (Singapur, Nueva Zelanda, Finlandia y Corea). <a href="http://dublincore.org/">http://dublincore.org/</a>
Otras iniciativas concluidas	GESMES. Servicio usado por el Banco Central Europeo para el intercambio de datos y metadatos, ahora incorporada al proyecto SDMX. AMRADS. Medidas empleadas para la investigación y desarrollo de estadísticas oficiales. MetaNet. Red creada para sintetizar y armonizar los desarrollos en metadatos estadísticos. COSMOS. Grupo de sistemas de metadatos para estadísticas oficiales. METAWARE. Sistema de metadatos estadísticos para almacenes de datos.

## Conociendo las estadísticas de México (COESME)

Este sistema tiene el propósito de dar a conocer a los usuarios una descripción o documentación general de los aspectos conceptuales, metodológicos y operativos que caracterizan las actividades de generación de información estadística del INEGI por medio de su clasificación en programas, proyectos y productos estadísticos.

## Data Warehouse (DWH)

Es un repositorio de información estadística para su consulta y análisis que permite a los usuarios el acceso a los datos de los proyectos estadísticos a cargo del INEGI. Este proyecto cuenta con un subsistema alineado con varios estándares internacionales, que integra toda la información necesaria para consultar e interpretar los datos, así como relacionar de manera efectiva cada uno con los metadatos correspondientes.

## Página del INEGI en Internet (PII)

A través de sus distintas secciones de consulta, ofrece a los usuarios un sitio para identificar y obtener información sobre aspectos metodológicos y conceptuales, documentos técnicos, así como el acceso a los microdatos de los proyectos estadísticos que lleva a cabo el Instituto.

## Implementación del DDI en el INEGI

### Sustento legal

En la *Ley del Sistema Nacional de Información Estadística y Geográfica*<sup>4</sup> (LSNIEG) se establece que la finalidad del Sistema Nacional de Información Estadística y Geográfica (SNIEG)<sup>5</sup> es la de suministrar a la sociedad y al Estado información de calidad,

4 [www.snieg.mx/contenidos/espanol/normatividad/marcojuridico/LSNIEG.pdf](http://www.snieg.mx/contenidos/espanol/normatividad/marcojuridico/LSNIEG.pdf)

5 [www.snieg.mx/](http://www.snieg.mx/)

pertinente, veraz y oportuna a efecto de coadyuvar al desarrollo nacional (art. 3); asimismo, indica que el SNIEG tendrá como objetivos producir información, difundirla oportunamente mediante mecanismos que faciliten su consulta, promover su conocimiento y uso, así como conservarla (art. 4).

De igual forma, la LSNIEG establece que el INEGI es la unidad central coordinadora del SNIEG y tiene como parte de sus funciones normar y coordinar las actividades que lleven a cabo las unidades del Estado en esta materia, tomando en cuenta los estándares nacionales e internacionales (art. 55); también, dispone que el Instituto deberá dar a conocer y conservar los metadatos o especificaciones concretas de la aplicación de las metodologías que se hubieren utilizado en la elaboración de la información (art. 88).

### Adopción del estándar

Bajo este marco jurídico, el INEGI promueve la adopción de la DDI como estándar para la documentación de proyectos de generación de información estadística básica, cuyo objetivo es desarrollar un sistema de documentación de microdatos de fácil acceso para los usuarios de diferentes niveles de especialización que permita el aprovechamiento óptimo de los datos.

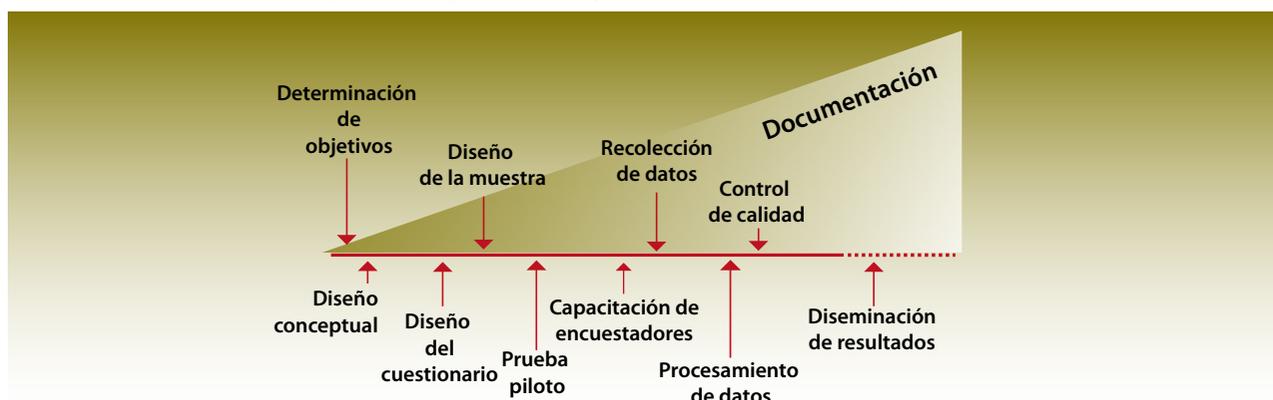
Con el propósito de cumplir estas funciones, se definieron las acciones estratégicas a seguir —que van alineadas con la iniciativa del Banco Mundial, el Consorcio de Estadísticas para el Desarrollo en el Siglo 21<sup>6</sup> (PARIS21) y el SNIEG— para llevar a cabo la implementación del uso de las Herramientas para la Gestión de Microdatos, las cuales facilitan la documentación, disseminación y preservación en forma estandarizada de los proyectos de generación de estadística básica a cargo tanto del INEGI como de las Unidades del Estado.

Actualmente, se cuenta con la metodología y logística que permite la documentación de los

6 [www.paris21.org/](http://www.paris21.org/)

Figura 2

## Proceso integrado de generación y documentación de proyectos estadísticos



proyectos de generación de información estadística básica de manera estandarizada; también, se cuenta con el sitio en Internet de la Red Nacional de Metadatos (RNM)<sup>7</sup> para la difusión de los proyectos documentados.

A mediano plazo, se pretende que todos los proyectos de generación de información de estadística básica del INEGI se documenten en el estándar DDI para ser incorporados a la RNM y que la documentación se realice de manera paralela a todas las etapas de generación de los proyectos, de tal manera que sea un proceso integral, eficiente y de alta calidad en sus resultados (ver figura 2).

### Red Nacional de Metadatos

Se trata de una herramienta que facilita la catalogación y difusión de los metadatos de proyectos de estadística básica de acuerdo con los estándares internacionales DDI y DCMI para la consulta e intercambio de datos y metadatos.

Su objetivo es ayudar a los usuarios, con diferentes niveles de especialización, a realizar búsquedas y consultas generales sobre las operaciones estadísticas o específicas a nivel de variables; asimismo, les da acceso controlado a los microdatos de

las operaciones estadísticas de su interés. Desde el punto de vista del productor de datos, la RNM es de gran utilidad para preservar los microdatos que se producen y difundirlos en forma estandarizada (ver figura 3).

El uso de la Red Nacional de Metadatos como medio de diseminación de metadatos permite, entre otras acciones:

- Buscar y encontrar los metadatos a nivel de variable de los proyectos estadísticos existentes para una determinada zona geográfica, referencia temporal, un tema determinado o, en general, con las características específicas que el usuario demanda.
- Utilizar con facilidad los microdatos mediante la descripción de todas las características técnicas relevantes de los mismos de un modo objetivo, amplio y completo, haciendo posible su explotación y ayudando a los usuarios tanto en la obtención de resultados como en su mantenimiento y actualización.
- Valorar la calidad del conjunto de datos.
- Comparar variables de distintos conjuntos de datos (ver figura 4), de modo que se puedan seleccionar cuáles cumplen los requisitos del usuario.
- Evitar la duplicidad de trabajo, difundiendo la información existente, su ubicación y disponibilidad.

<sup>7</sup> [www3.inegi.org.mx/rnm/index.php/catalog](http://www3.inegi.org.mx/rnm/index.php/catalog)

Figura 3



Figura 4

### Ejemplo de un ejercicio de comparación de variables entre tres levantamientos de un mismo proyecto



Al 15 de julio de 2013, en la RNM hay 41 proyectos estadísticos documentados, que son el resultado del esfuerzo institucional para lograr la implementación y operación de los esquemas de documentación de datos, así como de validación y publicación de metadatos.

A continuación se relacionan los proyectos y levantamientos que se encuentran disponibles para su consulta en Internet ([www3.inegi.org.mx/rnm/index.php/catalog](http://www3.inegi.org.mx/rnm/index.php/catalog)):

- Censos nacionales:
  1. Censo de Población y Vivienda 2010.
  2. Censos Económicos 2009.

3. Censo Nacional de Procuración de Justicia Estatal 2012.
4. Censo Nacional de Gobierno, Seguridad Pública y Sistema Penitenciario Estatales 2012.
5. Censo Nacional de Impartición de Justicia Estatal 2012.
6. Censo Nacional de Gobiernos Municipales y Delegacionales 2011.
- Encuestas en hogares:
  7. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2002.
  8. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2004.
  9. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2005.

10. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2006.
11. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2008.
12. ENIGH 2008. Módulo de Condiciones Socioeconómicas.
13. ENIGH 2008. Nueva construcción de variables.
14. Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2010.
15. ENIGH 2010. Módulo de Condiciones Socioeconómicas.
16. ENIGH 2010. Nueva construcción de variables.
17. Encuesta Nacional sobre Dinámica Demográfica 2009.
18. Encuesta Nacional sobre Uso del Tiempo 2009.
19. Encuesta sobre el Seguro Médico para una Nueva Generación 2009.
- Encuestas en establecimientos:
  20. Encuesta Mensual de Empresas Comerciales, 2004-2010.
  21. Encuesta Mensual de la Industria Manufacturera 2012.
  22. Encuesta Mensual de la Industria Manufacturera 2011.
  23. Encuesta Anual de la Industria Manufacturera, 2009-2010.
  24. Encuesta Mensual de Opinión Empresarial 2008.
  25. Encuesta Mensual de Opinión Empresarial 2009.
  26. Encuesta Mensual de Opinión Empresarial 2010.
  27. Encuesta Mensual de Servicios 2009.
  28. Encuesta Mensual de Servicios 2010.
  29. Encuesta Nacional de Empresas Constructoras 2010.
- Registros administrativos:
  30. Directorio Estadístico Nacional de Unidades Económicas, 2011.
  31. Estadísticas de finanzas públicas estatales y municipales, 2010.
  32. Estadísticas de transporte urbano de pasajeros, 2008-2012.
  33. Estadísticas del comercio exterior de México, 2010-2011.
  34. Estadística de sacrificio de ganado en rastro municipales, 2003-2011.
  35. Estadísticas del comercio exterior de México, 2009.
  36. Estadística de salud en establecimientos particulares, 2009-2011.
  37. Estadísticas sobre relaciones laborales de jurisdicción local, 2009-2011.

## Consideraciones finales

La producción de información estadística es un proceso muy complejo; por ello, resulta conveniente que sea documentada de forma detallada. Los metadatos estadísticos constituyen la información que permite que los datos estadísticos sean bien entendidos y aprovechados. En la actualidad, el INEGI cuenta con la metodología y la logística para llevar a cabo el proceso de documentación de microdatos bajo estándares internacionales; asimismo, cuenta con la RNM que cumple con el objetivo de difundir en un formato estandarizado los metadatos de referencia y de estructura de las operaciones estadísticas e, incluso, todos los documentos que dan sustento conceptual, metodológico y técnico a cada proyecto.

## Referencias

- INEGI. Dirección General de Estadística. *Manual de la estructura para metadatos estadísticos*. México, agosto del 2008.
- \_\_\_\_\_. Coordinación Técnica de la Vicepresidencia de Información Demográfica y Social. *Plan de documentación*. México, enero del 2011.
- \_\_\_\_\_. Coordinación Técnica de la Vicepresidencia de Información Demográfica y Social. *Guía rápida para la documentación de datos*. México, enero del 2011.
- \_\_\_\_\_. *Inventario de estándares estadísticos internacionales*. Disponible en: <http://mapserver.inegi.org.mx/estandares/Index.cfm>
- Ortega y Venzor, Alberto M. *Los estándares estadísticos internacionales y la armonización de las estadísticas nacionales*. México, INEGI, agosto del 2009.
- United Nations Economic Commission for Europe. *Statistical Metadata in a Corporate Context: A guide for managers*. Geneve, February, 2009.