

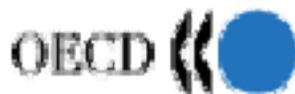
# La iniciativa del Programa Acelerado de Datos (PAD)

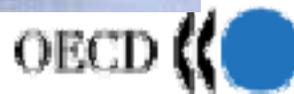
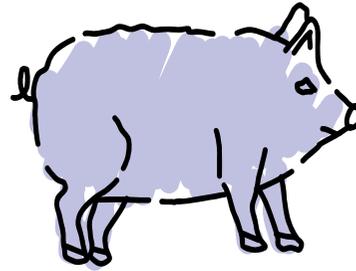
## TALLER **DANE** SENSIBILIZACIÓN SOBRE INSTRUMENTOS PARA COORDINACIÓN Y REGULACIÓN DEL SISTEMA ESTADÍSTICO NACIONAL

**Bogotá, COLOMBIA**

Septiembre 19, 2011

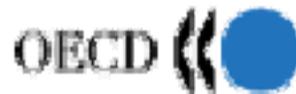
Alejandro Medina – Banco Mundial



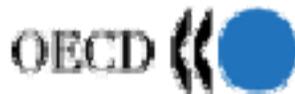
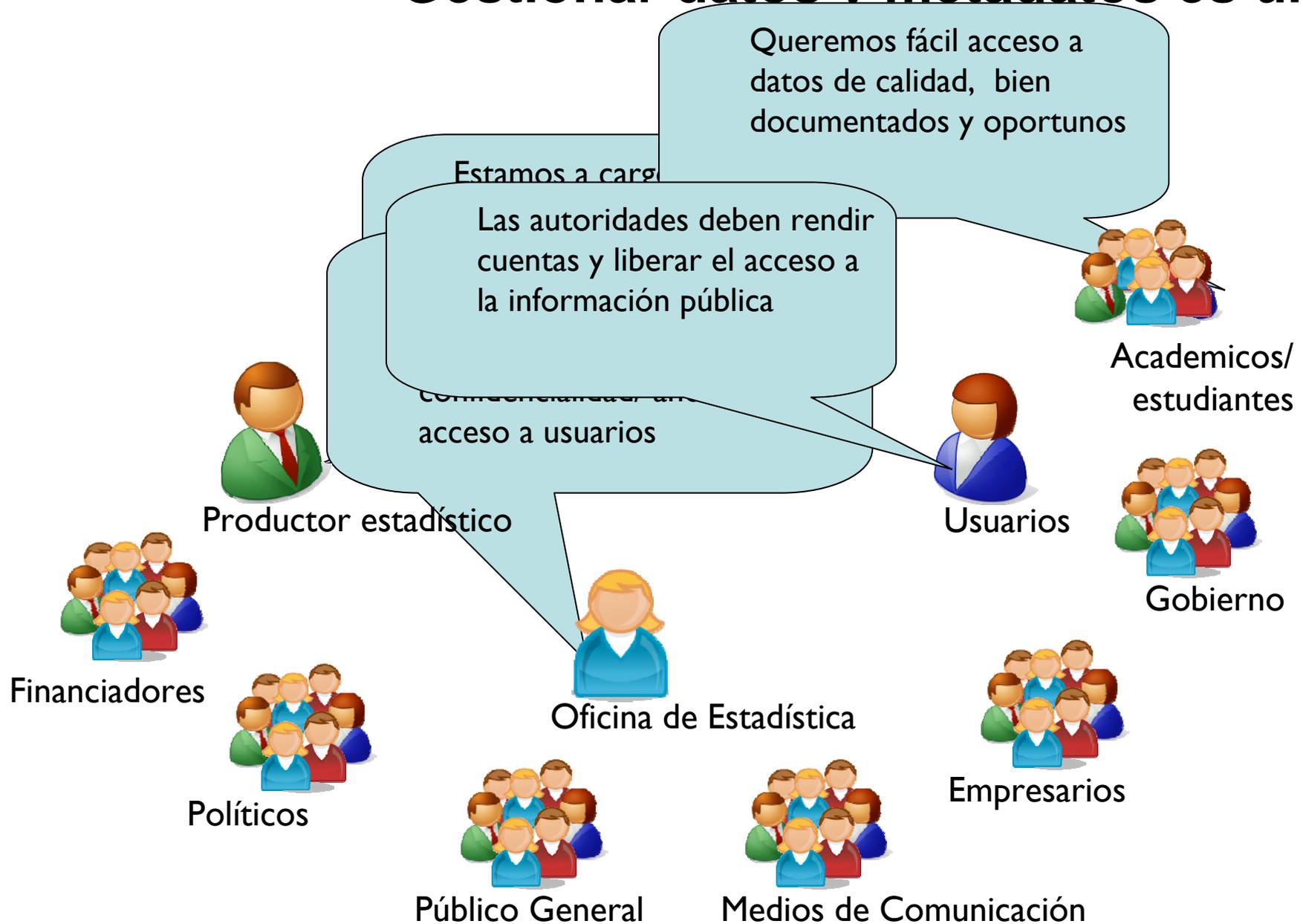


# Antecedentes

- Si ponemos en el centro a la **RELEVANCIA X INFORMACIÓN ESTADÍSTICA**, la respuesta depende también del **marco de referencia** desde donde se **vea**.
- Políticos, Estudiantes, Economistas, Funcionarios de Gobierno, ONG's, Multilaterales, Investigadores, Oficinas Nacionales de Estadística, Empresarios...etc.



# Gestionar datos y metadatos es un



# Antecedentes

## Motivación Programa Acelerado de Datos

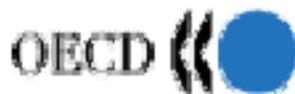
### Fuentes básicas de información estadística

- ***Microdatos*** fuente importante de **indicadores para monitorear** estrategias de combate a la **pobreza y evaluar *efectividad de políticas públicas***.
- ***Demanda creciente*** de investigadores, estudiantes, partidos políticos y medios de comunicación por **rendición de cuentas y liberación de datos**
- ***Pieza clave para fortalecer los sistemas nacionales y sectoriales de monitoreo y evaluación. SM&E***



# Antecedentes

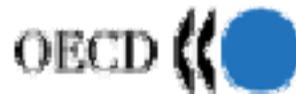
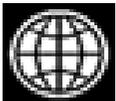
## DIAGNÓSTICO SOBRE LA OFERTA Y DEMANDA DE INFORMACIÓN ESTADÍSTICA EN AMÉRICA LATINA Y EL CARIBE



# Antecedentes

## Algunos problemas y obstáculos

- No hay **inventarios** completos y/o actualizados de operaciones estadísticas.
- No hay **estándares de calidad** en la producción, documentación y diseminación de datos.
- Pobre y altamente fragmentada **documentación** de operaciones estadísticas, sin uso de estándares
- No hay **políticas formales de diseminación de datos**

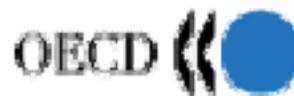


## QUEJAS FRECUENTES DE USUARIOS:

Visibilidad / *“no sabemos lo que ellos tienen”*  
Accesibilidad / *“ellos no quieren compartir los datos”*  
Confiabilidad / *“estos datos son un desorden”*  
Relevancia / *“esto no es realmente lo que necesito”*  
Uso / *“donde esta la documentación?”*



- **Falta de conexión entre decisores de política y productores estadísticos.** Se rompe el ciclo producción-usuario.



# Antecedentes

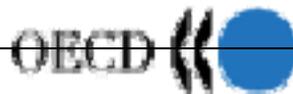
- La Iniciativa del **Programa Acelerado de Datos** apoya a mejorar la calidad de la **Oferta** de información estadística a través de la documentación y preservación de los microdatos (inventario, calidad, documentación, preservación) y apoya la **Demanda** a través de herramientas de diseminación para usuarios.



## Plan de Acción Estadístico Marrakech 2004.

### Cumbre Gestión de Resultados para el Desarrollo:

- Estrategias Nacionales para el Desarrollo Estadístico - países en desarrollo. **ENDE.**
- Ronda preparatoria **Censos 2010.**
- Desarrollar Red Internacional de Encuestas de Hogares **RIEH.**
- Mejoras urgentes para monitorear los **Objetivos de Desarrollo del Milenio.**



# Programa Acelerado de Datos (PAD)

## Objetivos y alcance



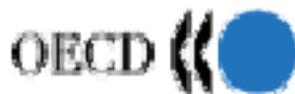
# Programa Acelerado de Datos (PAD)

Identifica 3 tareas:

**Tarea No 1: Documentación, diseminación y preservación** de microdatos (estándares internacionales DDI y DC)

**Tarea No 2: Análisis y evaluación de la calidad** de los microdatos existentes (armonización).

**Tarea No 3: Apoyo en las actividades de recolección** de datos (financiamiento a operaciones estadísticas).



# Programa Acelerado de Datos (PAD)

## Ventajas de la Documentación

- Preservación de datos y metadatos (**memoria institucional**).
- Documentación, archivo y diseminación apropiadas de datos promueven su **utilización**.
- Mayor y mejor uso de los datos incrementa su demanda y puede impulsar **aumento en el presupuesto** para recolección de datos.
- Si documentar **no es** una **parte integral** de la operación estadística, **requiere esfuerzos y recursos adicionales**.

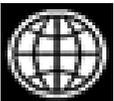


# Programa Acelerado de Datos (PAD)

Usualmente, documentar se considera una etapa **ex-post** y ajena al proceso de desarrollo de la operación estadística. Algo que “hay que hacer después que todo acaba”.

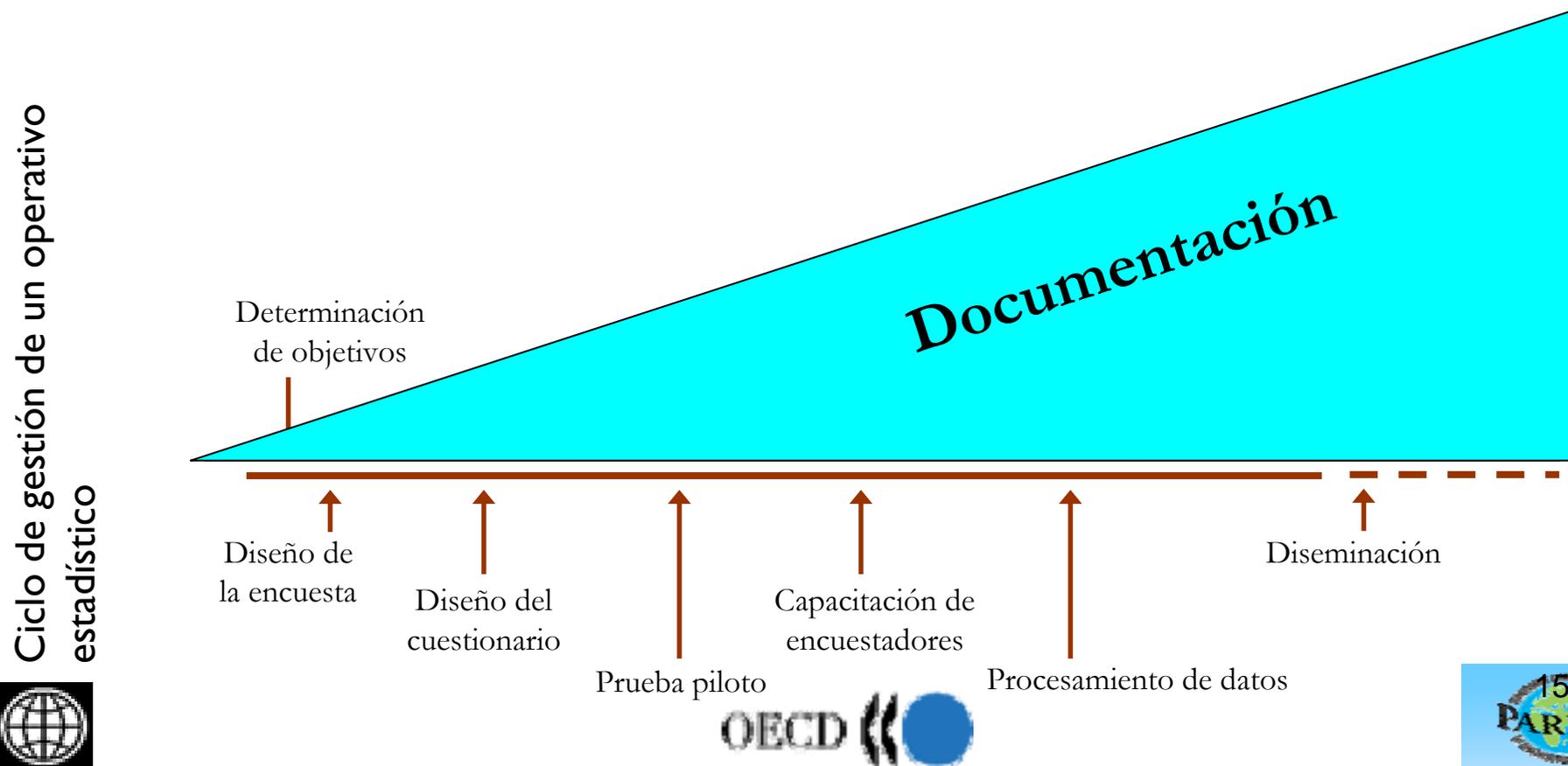
La mayor parte de la información relevante para documentar se genera en estas etapas del proceso, pero no se captura.

Documentación



# Programa Acelerado de Datos (PAD)

La documentación debe ser considerada como etapa integral y simultánea a la operación estadística.



# Programa Acelerado de Datos (PAD)

1	1	1	2	7
1	1	1	7	7
1	1	1	4	9
1	1	1	2	5
1	1	1	2	8
1	1	1	4	8
1	1	1	1	3
1	1	1	2	5
1	1	1	3	7
1	1	2	*	*
1	1	1	1	3
1	1	1	1	5
1	1	1	2	6
1	1	1	5	8
1	1	1	1	4

**Variable HC1B : Mother tongue**

**Overview**  
 Type: Continuous    Valid case(s): 4811 (4811)  
 Format: numeric    Invalid: 189 (0)  
 Width: 1    Minimum: 1  
 Decimal(s): 0    Maximum: 9  
 Range: 1-9

**File:** [hh](#)

**Universe**  
Households

**Source of information**  
Head of household or other responsible household member

**Categories**

Value	Category	Cases	weighted	
1	Language 1	3150	3132	65.1%
2	Language 2	867	893	18.6%
3	Language 3	195	186	3.9%
6	Other	598	599	12.5%
9	Missing	1	1	0.0%

**Literal question**  
What is the mother tongue/native language of the head of this household?

**Interviewer instructions**  
Circle the code corresponding to the answer given. Make sure to get the religion.



Source: Courtesy of Jostein Ryssevik, Nesstar Inc.

The bean example is taken from: A Manager's Introduction to Adobe eXtensible Metadata Platform, <http://www.adobe.com/products/xmp/pdfs/whitepaper.pdf>



# Ventajas

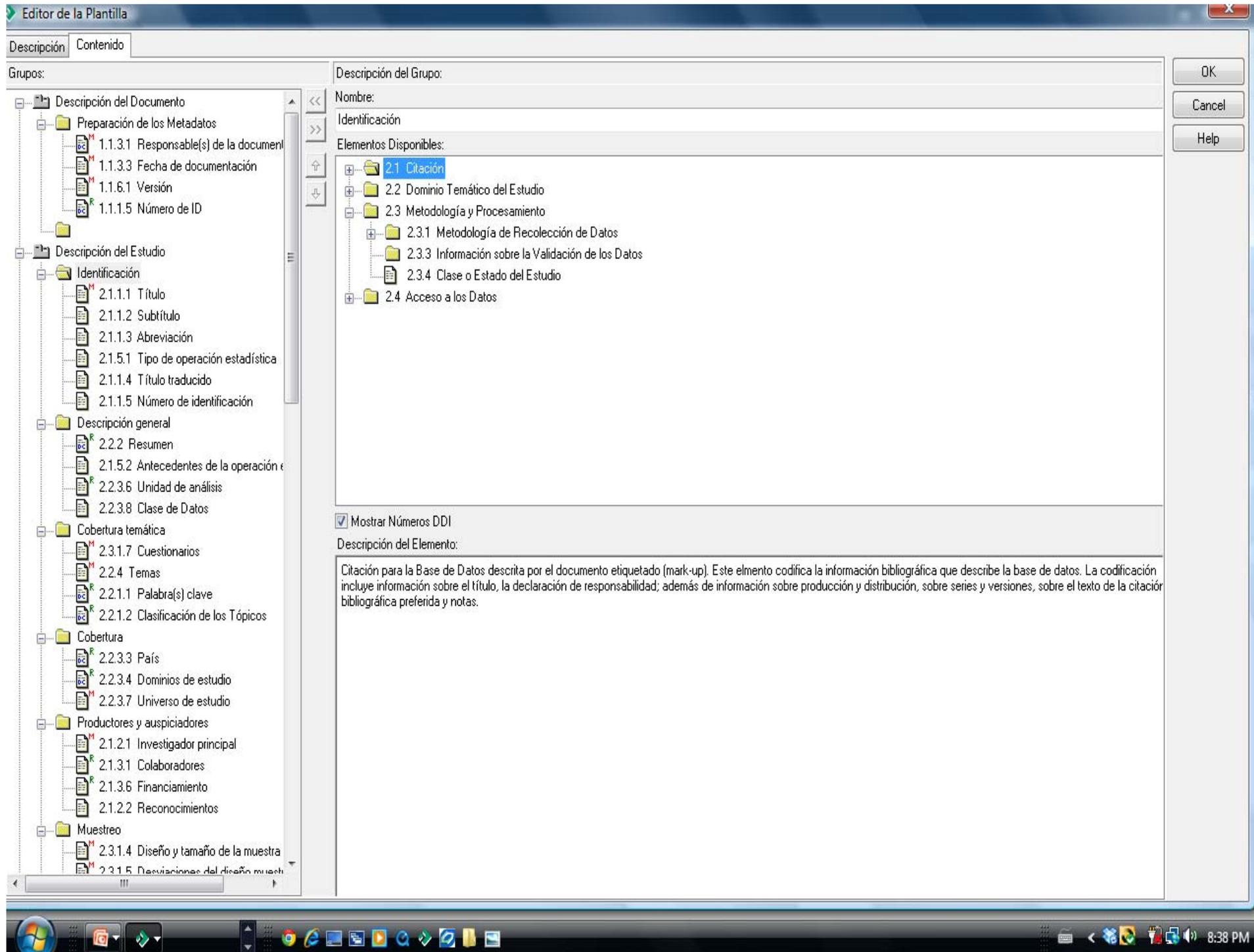
- Al **documentar** durante el **ciclo de gestión** se **detectan y previenen** errores de de codificación, se recupera información que de otro modo se perdería.
- **Mejoran procesos** de trabajo y mecanismos de coordinación.
- Se **reducen tiempos** en la publicación
- Se promueve la **publicación** de información, etc.



# Iniciativa de Documentación de Datos (IDD)

- Estándar internacional dedicado a documentar microdatos.
- Facilita la preservación e intercambio de documentación de bases de datos de las ciencias sociales.
- Basado en XML para explotar tecnología de Web.
- Diseñado originalmente por el Consorcio Inter-Universitario para la Investigación Política y Social (ICPSR) –500 miembros
- Hoy en día manejado por una alianza de instituciones líderes en documentación, preservación y diseminación de datos





Grupos:

- Descripción del Documento
  - Preparación de los Metadatos
    - 1.1.3.1 Responsable(s) de la document
    - 1.1.3.3 Fecha de documentación
    - 1.1.6.1 Versión
    - 1.1.1.5 Número de ID
- Descripción del Estudio
  - Identificación
    - 2.1.1.1 Título
    - 2.1.1.2 Subtítulo
    - 2.1.1.3 Abreviación
    - 2.1.5.1 Tipo de operación estadística
    - 2.1.1.4 Título traducido
    - 2.1.1.5 Número de identificación
  - Descripción general
    - 2.2.2 Resumen
    - 2.1.5.2 Antecedentes de la operación e
    - 2.2.3.6 Unidad de análisis
    - 2.2.3.8 Clase de Datos
  - Cobertura temática
    - 2.3.1.7 Cuestionarios
    - 2.2.4 Temas
    - 2.2.1.1 Palabra(s) clave
    - 2.2.1.2 Clasificación de los Tópicos
  - Cobertura
    - 2.2.3.3 País
    - 2.2.3.4 Dominios de estudio
    - 2.2.3.7 Universo de estudio
  - Productores y auspiciadores
    - 2.1.2.1 Investigador principal
    - 2.1.3.1 Colaboradores
    - 2.1.3.6 Financiamiento
    - 2.1.2.2 Reconocimientos
  - Muestreo
    - 2.3.1.4 Diseño y tamaño de la muestra
    - 2.3.1.5 Desviaciones del diseño muestr

Descripción del Grupo:

Nombre:

Identificación

Elementos Disponibles:

- 2.1 Citación
- 2.2 Dominio Temático del Estudio
- 2.3 Metodología y Procesamiento
  - 2.3.1 Metodología de Recolección de Datos
  - 2.3.3 Información sobre la Validación de los Datos
  - 2.3.4 Clase o Estado del Estudio
- 2.4 Acceso a los Datos

Mostrar Números DDI

Descripción del Elemento:

Citación para la Base de Datos descrita por el documento etiquetado (mark-up). Este elemento codifica la información bibliográfica que describe la base de datos. La codificación incluye información sobre el título, la declaración de responsabilidad; además de información sobre producción y distribución, sobre series y versiones, sobre el texto de la citación bibliográfica preferida y notas.

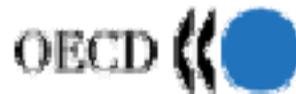
OK

Cancel

Help

# Herramientas de Gestión de Microdatos

- **Editor de Metadatos.**
  - Software desarrollado por el Servicio de Datos Noruego para las Ciencias Sociales
  - Trabajo de edición basado en los campos del DDI
  - Permite reciclar metadatos
  - Genera archivo en formato .nesstar que contiene base de datos (transformador de formatos); recursos externos manuales, cuestionarios, etc. y metadatos del DDI



Metadata Editor v1.1

Archivo Editar Documentación Variables Datos Publicando Herramientas Ayuda

Spanish

Proyectos:

- Mis Proyectos
  - HONDURAS
  - MEXICO
    - CUD\_GOLD
    - semefo2010
    - SEMEFO 2008\_1
    - ENIGH 2008
      - Descripción del Documento
      - Descripción de la operación estadística
        - Identificación
        - Descripción general
        - Cobertura temática
        - Cobertura
        - Productores y auspiciadores
        - Muestreo
          - Diseño y tamaño de la muestra**
          - Desviaciones del diseño muestral
          - Tasa de no respuesta
          - Factores de expansión
        - Recolección de datos
        - Procesamiento de datos
        - Evaluación de los datos
        - Acceso a la base de datos
        - Notas legales y derechos de autor
        - Contactos
      - Base de Datos
      - Grupo de Variables
      - Materiales de Referencia Externos
- RPUBLICA DOMINICANA SOLIDARIDAD

### Diseño y tamaño de la muestra

#### Diseño de la muestra

El diseño de la muestra para la ENIGH-2008 se caracteriza por ser probabilístico, en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población. A la vez el diseño es bietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar.

#### Marco de la encuesta

El marco de muestreo que se empleó para la ENIGH-2008 es el Marco Nacional de Viviendas 2002 del INEGI, construido a partir de la información cartográfica y demográfica que se obtuvo del XII Censo General de Población y Vivienda 2000. Este marco es en realidad una muestra maestra a partir de la cual se seleccionan las muestras para todas las encuestas en viviendas que realiza el INEGI. El diseño del presente marco es probabilístico, estratificado, unietápico y por conglomerados, estos últimos también se consideran unidades primarias de muestreo, pues es en ellos donde se seleccionan en una segunda etapa las viviendas que integran las muestras de las diferentes encuestas.

#### Formación de las unidades primarias de muestreo (UPM)

Las unidades primarias de muestreo están constituidas por agrupaciones de viviendas con características diferenciadas dependiendo del ámbito al que pertenecen, como se especifica a continuación:

a) En urbano alto

El tamaño mínimo de una UPM es de 80 viviendas habitadas y el máximo es de 160. Pueden estar formadas por:

- Una manzana.
- La unión de dos o más manzanas contiguas de la misma AGEB (Área geoestadística básica).
- La unión de dos o más manzanas contiguas de diferentes AGEB de la misma localidad.
- La unión de dos o más manzanas contiguas de diferentes localidades, que pertenezcan al mismo tamaño de localidad.

b) En complemento urbano

#### Descripción del elemento:

Información sobre el marco muestral y los procedimientos utilizados para seleccionar a los encuestados.

El tamaño de muestra seleccionado y su distribución también debe ser indicado.

8:24 PM

# Herramientas de Gestión de Microdatos

- Nesstar Explorer
  - Es un navegador, es un medio de publicación. No permite edición
  - Incluye un transformador de formatos a la base de datos, se puede exportar la base de datos en los formatos más comunes dbf, stata, spss, etc.



Contenido del Estudio:

- Descripción del Documento
- Descripción de la operación estadística
- Base de Datos
- Grupo de Variables
- Materiales de Referencia Externos

### Descripción del Documento

#### Responsable(s) de la documentación

Name

Mtra. Ana Laura Pineda Manríquez

#### Fecha de documentación

2009-01-23

#### Versión

Versión 3 (23 de enero de 2009). Esta versión contiene toda la base de datos de la ENIGH 2006.

#### Número de ID

001-MEX-INEGI-ENIGH-2006

**Exportar Base de Datos**

Exportar en formato:  
dBase (\*.dbf)

Exportar al Directorio:  
C:\Users\al\Desktop

Crear Nuevo Sub Directorio

Nombre del Subdirectorio:

Base de Datos:

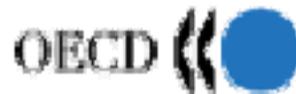
- Hogares
- Poblacion
- Ingresos
- Gastos
- Erogaciones
- Nomonetario
- Concentrado

OK Cancel

# Herramientas de Gestión de Microdatos

- **Cd Rom Builder**

- Genera archivos en formato HTML que pueden publicarse en páginas Web
- Genera CDs automáticos
- Permite descargar las bases de datos en formatos comunes



SEMEFO CUD



**IHSN Microdata Management Toolkit**  
Encuesta Nacional de Ingresos y Gastos de los Hogares 2008

- SEMEFO
  - Introducción
  - Descripción general
  - Información técnica
  - Bases de datos
    - Política de acceso a la informac
    - Archivos de datos
    - Diccionario de variables
      - Grupo Características de la
      - Grupo Residentes e identific
      - Grupo Características soci**
      - Grupo Acceso a la alimenta
      - Grupo Equipamiento del hoq
      - Grupo Tiempo de traslado a
      - Grupo Discapacidad en los
      - Grupo Ocupación de los int
      - Grupo Cohesión social
      - Grupo Uso del tiempo
      - Grupo Salud
      - Grupo Negocios del hogar
  - Cuadros e informes
  - Otros materiales

## Grupo Características sociodemográficas

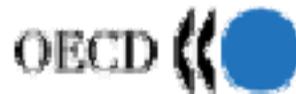
[Presione aquí para acceder/exportar al archivo\(s\) de datos.](#)  
Para abrir este archivo se necesita el software gratuito [Nesstar Explorer](#) software. El Explorador NESSTAR permite visualizar los archivos de datos y exportarlos a diferentes formatos estadísticos.

### Variable(s)

#	Nombre	Etiqueta	Tipo	Formato	Válido	Inválido	Pregunta
1	<a href="#">parentesco</a>	<a href="#">Parentesco</a>	Discrete	numeric 4.0	118927	0	¿Qué parentesco tiene ... (NOMBRE) con el jefe o jefa del hogar?
2	<a href="#">sexo</a>	<a href="#">Sexo</a>	Discrete	numeric 3.0	118927	0	... (NOMBRE) es...?
3	<a href="#">edad</a>	<a href="#">Edad</a>	Continuous	numeric 3.0	118927	0	¿Cuántos años cumplidos tiene ... (NOMBRE)?
4	<a href="#">edo_salud</a>	<a href="#">Percepción del estado de salud</a>	Discrete	numeric 3.0	118927	0	En general ¿como calificaría el estado de salud de ... (NOMBRE) hoy...?
5	<a href="#">lengua6</a>	<a href="#">Lengua indígena</a>	Continuous	numeric 3.0	112732	6195	¿... (NOMBRE) habla alguna lengua indígena o dialecto?
6	<a href="#">lengua7</a>	<a href="#">Que lengua indígena</a>	Discrete	character 4	10116	0	¿... (NOMBRE) que lengua indígena o dialecto habla?
7	<a href="#">lengua8</a>	<a href="#">Español</a>	Continuous	numeric 3.0	10116	108811	¿... (NOMBRE) también habla español?
8	<a href="#">alfabe</a>	<a href="#">Alfabetismo</a>	Continuous	numeric 3.0	112732	6195	¿... (NOMBRE) sabe leer y escribir un recado?
9	<a href="#">asis_esc</a>	<a href="#">Asistencia a la escuela</a>	Continuous	numeric 3.0	112732	6195	¿... (NOMBRE) asiste actualmente a la escuela?
10	<a href="#">nivel</a>	<a href="#">Nivel escolar al que asiste</a>	Continuous	numeric 3.0	37013	81914	¿Cuál es el año o grado al que asiste... (NOMBRE)?
11	<a href="#">orado</a>	<a href="#">Grado escolar al que asiste</a>	Continuous	numeric 3.0	37013	81914	¿Cuál es el año o orado al que asiste... (NOMBRE)?

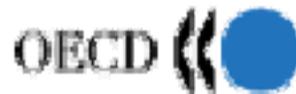
# Progreso en América Latina

- **Oficinas Nacionales de Estadística, Ministerios y Nivel Subnacional en 16 países de la región. Subregional. Oficial en SICA-CENTROESTAT, aprobado por el Comité Andino de Naciones. CAN**
- **Mas de 1,500 operaciones estadísticas documentadas. Portales en línea**



# Agenda en América Latina

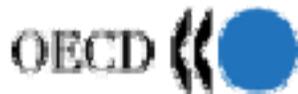
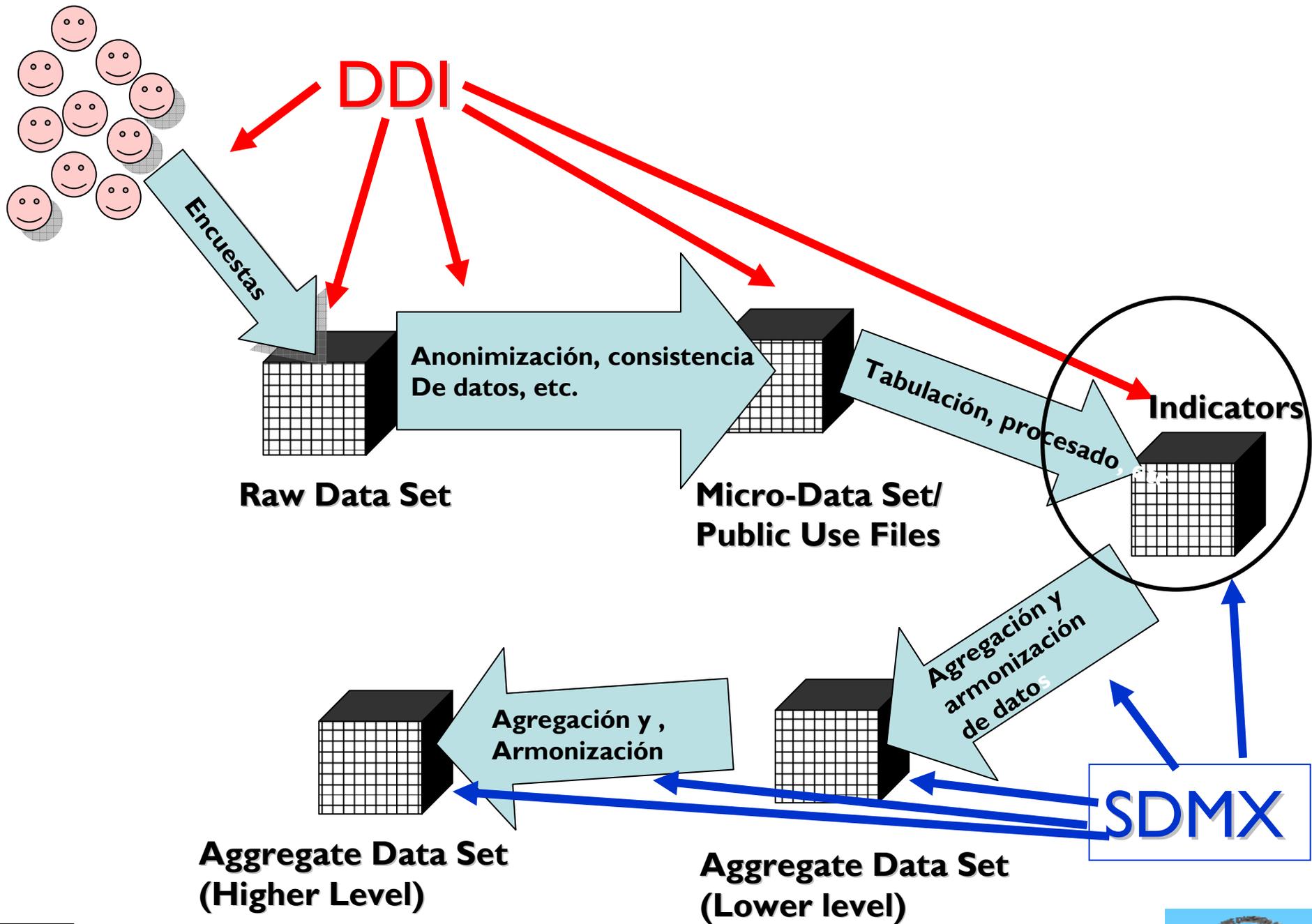
- **Avanzar con la expansión del DDI en los sistemas estadísticos nacionales**
  - ONE´s se encuentran expandiendo el DDI como rectores del SEN
  - Verificación de la calidad de la documentación
- **Avanzar en aspectos relativos a la armonización de preguntas/variables**
  - -Iniciativas relativas a la creación de banco de preguntas y bancos de cuestionarios

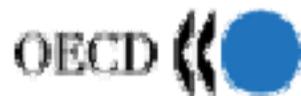


# Agenda en América Latina

- **Avanzar con aspectos de diseminación**
  - Técnicas avanzadas para anonimización de microdatos
  - Políticas y reglamentos para la diseminación de microdatos
  - Publicación de catálogos ANDA
  - **Otros**
    - Integración del DDI al estándar SDMX (para estadística agregada)







# Antecedentes

## *Microdatos*

- Datos acerca de objetos individuales (personas, establecimientos comerciales, transacciones, etc,) que tienen atributos concretos, tales como: “nombre”, “dirección”, “edad”, “ingreso”, etc.
- En encuestas y censos son registros a nivel de cada individuo entrevistado.

