



**DANE**  
Para tomar decisiones



**TODOS POR UN  
NUEVO PAÍS**  
PAZ EQUIDAD EDUCACIÓN

11

# **Lineamientos para la documentación de metadatos a partir de los estándares DDI y Dublin Core**

*Dirección de Regulación, Planeación,  
Estandarización y Normalización*

**DIRPEN**

**Agosto de 2014**

CÓDIGO: DSO-050-LIN-01  
VERSIÓN 1  
FECHA: 29-08-2014

## **Directora Técnica Dirección de Regulación, Planificación, Estandarización y Normalización (DIRPEN)**

Nelcy Araque García

## **Coordinación Regulación Estadística**

Fredy Yahirs Rodríguez Galvis

## **Líder Conceptualización y Diseño**

Grace Andrea Torres Pineda

## **Equipo técnico**

Paola Fernanda Medina Tovar

Marly Johana Téllez López

Diana Cristina Prieto Peña

Rafael Humberto Zorro

## **Revisión de coherencia**

Sonia Marcela Naranjo Morales

Agradecemos a las Coordinaciones de la DIRPEN y a las Direcciones Técnicas del DANE,  
quienes contribuyeron con sus aportes en la elaboración de este documento

# CONTENIDO

**PRESENTACIÓN ..... 6**

**INTRODUCCIÓN ..... 7**

**1. OBJETIVOS ..... 8**

**2. ALCANCE..... 9**

**3. CONCEPTOS BÁSICOS..... 10**

**4. IDENTIFICACIÓN DE ACTORES ..... 11**

**5. ESTÁNDARES DDI Y DUBLIN CORE ..... 12**

**6. PROCESO DE DOCUMENTACIÓN..... 13**

**BIBLIOGRAFÍA ..... 47**

# INDÍCE DE TABLAS

**Tabla 1. Colaboradores ..... 28**

**Tabla 2. Financiamiento ..... 29**

**Tabla 3. Periodo de recolección ..... 32**

**Tabla 4. Periodo de referencia..... 34**

**Tabla 5. Recolección de datos..... 34**

**Tabla 6. Institución propietaria ..... 35**

## INDÍCE DE GRÁFICAS

<b>Gráfica 1. Descripción del documento.....</b>	<b>14</b>
<b>Gráfica 2. Responsables de la documentación .....</b>	<b>15</b>
<b>Gráfica 3. Identificador único .....</b>	<b>16</b>
<b>Gráfica 4. Descripción de la operación estadística .....</b>	<b>16</b>
<b>Gráfica 5. Identificación .....</b>	<b>17</b>
<b>Gráfica 6. Título .....</b>	<b>17</b>
<b>Gráfica 7. Subtítulo .....</b>	<b>17</b>
<b>Gráfica 8. Abreviación .....</b>	<b>18</b>
<b>Gráfica 9. Tipo de estudio .....</b>	<b>18</b>
<b>Gráfica 10. Identificador de la operación estadística .....</b>	<b>19</b>
<b>Gráfica 11. Descripción general.....</b>	<b>20</b>
<b>Gráfica 12. Clase de datos .....</b>	<b>22</b>
<b>Gráfica 13. Cobertura temática .....</b>	<b>23</b>
<b>Gráfica 14. Palabras claves .....</b>	<b>25</b>
<b>Gráfica 15. Clasificación de los tópicos .....</b>	<b>25</b>
<b>Gráfica 16. Cobertura geográfica.....</b>	<b>26</b>
<b>Gráfica 17. País .....</b>	<b>26</b>
<b>Gráfica 18. Productores y auspiciadores .....</b>	<b>27</b>
<b>Gráfica 19. Muestreo .....</b>	<b>29</b>
<b>Gráfica 20. Recolección de datos.....</b>	<b>31</b>
<b>Gráfica 21. Método de recolección .....</b>	<b>32</b>
<b>Gráfica 22. Notas sobre la recolección de los datos.....</b>	<b>33</b>
<b>Gráfica 23. Acceso a las bases de datos.....</b>	<b>35</b>
<b>Gráfica 24. Contactos.....</b>	<b>36</b>
<b>Gráfica 25. Descripción del archivo .....</b>	<b>37</b>
<b>Gráfica 26. Listado de variables.....</b>	<b>39</b>
<b>Gráfica 27. Documentación .....</b>	<b>41</b>
<b>Gráfica 28. Preguntas .....</b>	<b>42</b>
<b>Gráfica 29. Imputación y derivación.....</b>	<b>43</b>
<b>Gráfica 30. Otros .....</b>	<b>44</b>
<b>Gráfica 31. Materiales de referencia externos .....</b>	<b>45</b>
<b>Gráfica 32. Ejemplo de presentación de material de referencia externo...</b>	<b>46</b>
<b>Gráfica 33. Ejemplo de presentación de los nombres de los documentos .</b>	<b>46</b>
<b>Gráfica 34. Tipo de material de referencia .....</b>	<b>47</b>

## PRESENTACIÓN

El Departamento Administrativo Nacional de Estadística (DANE), como coordinador del Sistema Estadístico Nacional (SEN) y en el marco del proyecto de Planificación y Armonización Estadística, trabaja por el fortalecimiento y consolidación del SEN mediante los siguientes procesos: la producción de estadísticas estratégicas; la generación, adaptación, adopción y difusión de estándares; la consolidación y armonización de la información estadística y la articulación de instrumentos, actores, iniciativas y productos. Estas acciones tienen como fin mejorar la calidad de la información estadística estratégica, su disponibilidad, oportunidad y accesibilidad para responder a la gran demanda que se tiene de ella.

Consciente de la necesidad y obligación de brindar a los usuarios mejores productos, el DANE desarrolló una serie de lineamientos que contribuyen a la visualización y el entendimiento del proceso estadístico. Allí se presentan de manera estándar, completa y de fácil lectura las principales características técnicas de los procesos y subprocesos de cada investigación, lo que permite su análisis, control, replicabilidad y evaluación.

Esta serie de documentos favorecen la transparencia, confianza y credibilidad de la calidad técnica de la institución para un mejor entendimiento, comprensión y aprovechamiento de la información estadística. Tal información es producida bajo los principios de coherencia, comparabilidad, integralidad y calidad de las estadísticas.

## INTRODUCCIÓN

El DANE consciente de la necesidad de brindar mayor claridad, transparencia y confianza técnica a los usuarios, en cuanto al proceso de las operaciones estadísticas que se desarrollan en la entidad y en aquellas instituciones que forman parte del SEN, adoptó en el 2009 una iniciativa basada en un estándar estadístico internacional denominado Programa Acelerado de Datos (PAD). Este programa tiene como propósito documentar, difundir y preservar metadatos y microdatos conforme a los estándares y prácticas internacionales por medio de la herramienta Nesstar Publisher. De igual manera, permite el desarrollo de parámetros para la presentación de los documentos técnicos y metodológicos utilizados por cada operación estadística.

A lo largo del tiempo, se ha incrementado la necesidad de contar con sistemas de información que capturen y compilen la información, presentándola de una forma amigable al usuario, y orientándola según los principios de accesibilidad, transparencia, confiabilidad y oportunidad. Es así como surge la necesidad de documentar con metadatos y microdatos estructurados y estandarizados, siendo de gran importancia la organización y el mantenimiento de la memoria institucional para la preservación de la información y los datos que la acompañan.

Esta guía proporciona, tanto a entidades públicas, privadas como académicas, los lineamientos necesarios para la documentación de metadatos y microdatos mediante los estándares DDI y Dublin Core, con el fin de facilitar su implementación y tener un recurso estandarizado con un vocabulario común.

## 1. OBJETIVOS

### Objetivo general

Proporcionar lineamientos para la documentación de las operaciones estadísticas producidas por las entidades pertenecientes al SEN mediante el uso de los estándares DDI y Dublin Core. Esto teniendo en cuenta tanto el contexto, la calidad, la condición, las características de los datos, los microdatos y el material relacionado, para la utilización de un estándar estadístico internacional y así obtener metadatos armonizados y estandarizados.

### Objetivos específicos

- Promover el uso de estándares estadísticos internacionales para la documentación de operaciones estadísticas en el SEN.
- Definir los pasos que se deben seguir al momento de realizar la documentación usando los Estándares DDI y Dublin Core.
- Establecer la información requerida para realizar la documentación de una operación estadística.



## 2. ALCANCE

La documentación de metadatos se apoya en la estandarización de documentos metodológicos, en donde se precisan los contenidos técnicos acordes con los procesos de la actividad estadística. Con esto se busca facilitar el cumplimiento de los principios fundamentales para el uso y la integración de las estadísticas y así contribuir con la comparación nacional, regional e internacional. Este proceso se aplica a todas las entidades productoras de estadísticas pertenecientes al SEN y debe ser desarrollada por sus funcionarios.

Este documento proporciona los lineamientos necesarios para la documentación de los metadatos en el contexto de su creación, utilización y conservación, teniendo en cuenta que los objetivos de los Estándares DDI y Dublin Core son:

- Brindar un mayor conocimiento de las operaciones estadísticas producidas en el SEN.
- Contribuir a la memoria institucional.
- Brindar a las personas que toman decisiones sobre política pública y a la sociedad en general, las características de las operaciones estadísticas que se producen en el país.
- Contribuir en la coordinación entre los diferentes productores y usuarios.

### 3. CONCEPTOS BÁSICOS

#### **Estándar estadístico:**

Proporciona, a nivel estadístico, un conjunto completo de directrices para las encuestas y fuentes administrativas recogiendo información sobre un tema en particular. El uso de estándares estadísticos permite repetir la recolección de estadísticas sobre una base constante. También permiten la integración de datos a lo largo del tiempo y entre diferentes fuentes de datos, lo que permite el uso de datos más allá del objetivo inmediato para el que se haya producido. Las normas también reducen las necesidades de recursos relacionados con muchos aspectos de desarrollo y mantenimiento de la encuesta (OCDE).

#### **Estándar estadístico internacional:**

Conjunto de guías estadísticas y recomendaciones internacionales que han sido desarrolladas por organizaciones internacionales con el trabajo de agencias nacionales. Los estándares cubren casi cada campo del ámbito estadístico; desde el diseño, la recolección de datos, el procesamiento y la difusión. Tales estándares también incluyen clasificaciones estadísticas internacionales (OCDE).

#### **Metadatos:**

Información necesaria para el uso e interpretación de las estadísticas. Los metadatos describen la conceptualización, calidad, generación, cálculo y características de un conjunto de datos estadísticos (DANE, 2012e).

#### **Microdatos:**

Datos sobre las características de las unidades de estudio de una población (individuos, hogares, establecimientos, entre otros), que constituyen una unidad de información en una base de datos y que son recogidos por medio de una operación estadística (DANE, 2012e).

#### **Operación estadística:**

Conjunto de procesos y actividades que, partiendo de la recolección sistemática de datos, conduce a la producción de resultados agregados (DANE, 2012e).

## 4. IDENTIFICACIÓN DE ACTORES

En el proceso de documentación de metadatos interviene el equipo temático de la operación u operaciones estadísticas que se documentan de las diferentes entidades pertenecientes al SEN, es decir, aquella persona que tenga todo el conocimiento sobre la actividad estadística, una vez se realiza la documentación, un delegado de la entidad deberá realizar la validación y aprobación de toda la información consignada en el *Nesstar Publisher*.

## 5. ESTÁNDARES DDI Y DUBLIN CORE

El PAD cuenta como herramienta básica con el *Nesstar Publisher* que es un editor para la documentación y preparación de los metadatos y los datos para su publicación en el catálogo en línea llamado Archivo Nacional de Datos (ANDA), que fue desarrollado por el Grupo de Datos para la Red Internacional de Encuesta de Hogares (IHSN).

El *Nesstar* está diseñado para promover la adopción de estándares internacionales para la documentación, la difusión y la preservación de los metadatos y microdatos. Esta herramienta permite un proceso de documentación que contiene dos estándares internacionales con el fin de contar con información organizada. Estos son el DDI y el Dublin Core *Metadata Initiative*.

- **Iniciativa de Documentación de Datos (DDI):** es un esfuerzo internacional para establecer un estándar basado en XML para la documentación de los microdatos. Su objetivo es proporcionar un instrumento simple para registrar y comunicar todas las características sobresalientes de los mismos. Estos aspectos son utilizados en la plantilla de estudios en el *Nesstar Publisher*.
- **Dublin Core Metadata Initiative (DCMI):** es un conjunto de elementos para la descripción de recursos digitales. Esta iniciativa es particularmente útil para describir los recursos relacionados a los microdatos, como son: cuestionarios, informes, manuales, scripts y programas de procesamiento de datos, etc. Estos elementos son identificados en la plantilla de descripción del material de referencia en el *Nesstar Publisher*.

## 6. PROCESO DE DOCUMENTACIÓN

Este inicia con la clasificación de la información que será documentada y luego ingresada en el Nesstar Publisher, en donde se encuentran configurados los estándares DDI y DCMI para iniciar el proceso de documentación. Es importante contar con la siguiente información respecto a la operación estadística a documentar: su metodología, las bases de datos y el material de referencia usados.

### Metodología

Se debe disponer de la descripción de los elementos necesarios para este documento. Puede ser consultada en la Guía para la elaboración de documentos metodológicos estándar de las operaciones estadísticas<sup>1</sup>.

### Base de datos

Se debe contar con la siguiente información:

- **Bases de datos (con o sin microdatos):** contiene las variables y registros de una operación estadística. Debe surtir un proceso de anonimización de las bases de datos con el fin de subirlas a un estándar estadístico internacional y difundirlos para la consulta de los usuarios.
- **Diccionario de datos:** contiene información de la base de datos como puede ser el nombre de la tabla o archivo de donde viene la variable, el nombre del campo, la descripción del campo, el tipo y medida del dato y la longitud del campo.
- **Manual de diligenciamiento:** contiene la descripción de cada una de las preguntas del formulario y lo que debe contestarse en cada una de ellas, lo cual permite la documentación de las variables que harán parte de la base de datos de la operación estadística.
- **Manual de validación y consistencia:** contiene información de la base de datos como el nombre de la tabla o archivo de donde viene la variable, el nombre del campo, la descripción del campo, el tipo de dato, la longitud del campo y adicionalmente, las reglas de validación y consistencia de la base de datos.

<sup>1</sup> Disponible en: [http://www.dane.gov.co/files/planificacion/fortalecimiento/cuadernillo/Elaboracion\\_documentos\\_metodologicos.pdf](http://www.dane.gov.co/files/planificacion/fortalecimiento/cuadernillo/Elaboracion_documentos_metodologicos.pdf)

## Material de referencia

Corresponde a aquellos documentos que constituyen el apoyo temático del proceso estadístico de la operación. Aquí se incluyen:

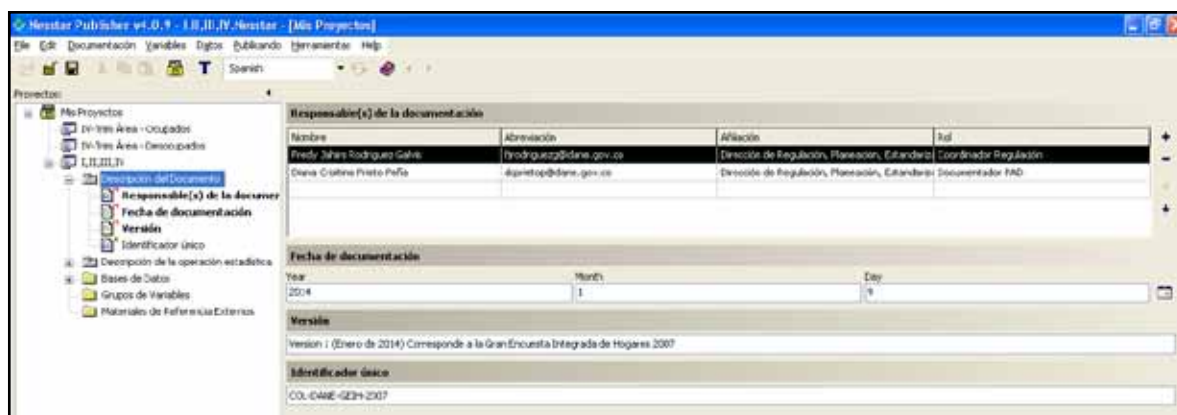
- Manuales.
- Documentos de diseño.
- Especificaciones de indicadores.
- Guías.
- Instructivos.
- Presentaciones.

Luego de clasificar los documentos se procede a realizar la documentación en el Nesstar Publisher, teniendo en cuenta que abarca cuatro secciones (descripción del documento, descripción de la operación estadística, base de datos y material de referencia). A continuación se identifica cada sección.

### 6.1. Descripción del documento

La operación estadística no siempre es documentada y difundida por el mismo organismo que produjo los datos. Es importante proveer en el metadato la información sobre las personas y/o entidades que intervinieron en la operación y durante el proceso de su documentación. En esta sección se incluyen los siguientes campos:

Gráfica 1. Descripción del documento



**Responsable(s) de la documentación**

Nombre	Abreviación	Afiliación	rol
Fredy Jairo Rodríguez Gahré	frodruiguez@dane.gov.co	Dirección de Regulación, Planeación, Estándares	Coordinador Regulación
Diana Cristina Prieto Peña	dprieto@dane.gov.co	Dirección de Regulación, Planeación, Estándares	Documentador RAD

**Fecha de documentación**

Year: 2014 Month: 1 Day: 9

**Versión**

Version 1 (Enero de 2014) Corresponde a la Gran Encuesta Integrada de Hogares 2007

**Identificador único**

COL-048E-G2H-2007

Fuente. Elaboración propia tomada del Nesstar Publisher

Responsable(s) de la documentación			
Nombre	Abreviación	Afiliación	Rol
Fecha de documentación			
Year	Month	Day	
Versión			
Identificador único			

Fuente. Elaboración propia tomada del Nesstar Publisher

**Responsables de la documentación:** contiene el nombre, afiliación y rol de los individuos y organizaciones involucradas en la creación de la documentación de la operación estadística (no son necesariamente los productores).

Gráfica 2. Responsables de la documentación

Responsable(s) de la documentación			
Nombre	Abreviación	Afiliación	Rol
Fredy Jairo Rodríguez Galvis	frodiguez@ dane.gov.co	Dirección de Regulación, Planeación, Estandarización	Coordinador Regulación
Diana Cristina Prieto Peña	dprieto@ dane.gov.co	Dirección de Regulación, Planeación, Estandarización	Documentador PAD

Fuente. Elaboración propia tomada del Nesstar Publisher

**Fecha de documentación:** corresponde a la fecha en que la documentación de la operación estadística fue elaborada. Esta casilla debe ser actualizada cada vez que el documento es revisado y modificado.

**Versión:** esta opción permite hacer un seguimiento a la versión del documento, lo que facilita a los usuarios la capacidad de determinar si cuentan con la versión más reciente.

También incluye un listado de cambios hechos en cada revisión. Esta información puede ayudar a los usuarios a determinar si los errores en versiones anteriores fueron la fuente de errores en el análisis.

Ej.: Versión 1 (Enero de 2011)  
Versión 2 (Febrero de 2011). Esta versión reemplaza la versión 1 (Enero de 2011) ya que incluye una descripción más detallada de la muestra.

**Identificador único:** corresponde al código único de identificación del documento en un archivo. Define y usa el esquema país-productor-operación, estadística-año en donde:

- País: corresponde a la abreviación de 3 letras tipo ISO.
- Productor: corresponde a la abreviación de la agencia productora.
- Operación estadística: corresponde a la abreviación de la operación.
- Año: corresponde en el que realizó el documento de metadatos.

Gráfica 3. Identificador único

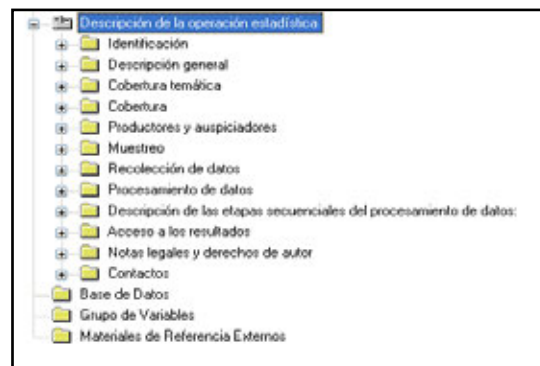
Identificador único
COL-DANE-GEIH-2012
Descripción del elemento:
Definir un identificador numérico o alfanumérico único para el metadato. El campo del identificador sólo debe contener letras, números, guiones al medio (-) y puntos.
Usar el siguiente esquema: País-Productor-Operación estadística-Año, donde:
- País: abreviación de 3 letras tipo ISO.
- Productor: abreviación de la agencia productora.
- Operación estadística: abreviación que identifica la operación estadística.
- Año: año al que corresponde la información documentada en el metadato (año del período de referencia).

Fuente. Elaboración propia tomada del Nesstar Publisher

## 6.2. Descripción de la operación estadística

Esta consta de la información general sobre la operación estadística; incluye la información sobre cómo debe ser citada, quién recolectó, compiló y distribuyó los datos, el contenido de los datos, los métodos de recolección de datos y procesamiento, etc.

Gráfica 4. Descripción de la operación estadística



Fuente. Elaboración propia tomada del Nesstar Publisher



Dentro de la descripción de la operación estadística se deben documentar los siguientes aspectos:

**Identificación:** los elementos de este grupo se refieren a la operación estadística así:

Gráfica 5. Identificación

Título	
Subtítulo	
Abreviación	
Tipo de estudio	1-2-3 Encuesta, fase 1 [hh/123-1]
Título traducido	
Identificador de la operación estadística	

Fuente. Elaboración propia tomada del Nesstar Publisher

**Título:** hace referencia al nombre completo de la operación estadística, incluyendo el año de ejecución (si es relevante). Las abreviaciones no deben ser utilizadas en esta casilla. El nombre de la operación estadística debe ir en mayúscula y con el año.

Gráfica 6. Título

Título	Gran Encuesta Integrada de Hogares 2012
--------	---

Fuente. Elaboración propia tomada del Nesstar Publisher

**Subtítulo:** corresponde al título secundario de la operación estadística. Generalmente la etapa, el trimestre u otra información similar.

Gráfica 7. Subtítulo

Subtítulo	
-----------	--

Fuente. Elaboración propia tomada del Nesstar Publisher

**Abreviación:** corresponde a la abreviación o el acrónimo conocido más utilizado para referirse a la operación estadística. Debe ir en mayúscula y sin el año.

Gráfica 8. Abreviación

Abreviación

GEIH

Descripción del elemento:

Definir la abreviación o siglas utilizada para referirse a la operación estadística.

Ejemplos:

- IPC
- ENDS
- ENSIN

Fuente. Elaboración propia tomada del Nesstar Publisher

**Tipo de estudio:** según la temática y el método de recolección o acopio se selecciona el tipo de estudio que se presenta en la lista desplegable. La plantilla utiliza un vocabulario controlado para este elemento. Si ninguna de las opciones listadas es la apropiada, incorpore la categoría correspondiente con una nueva entrada en el editor.

Gráfica 9. Tipo de estudio

Tipo de estudio

Otras encuestas a hogares [hh/oth]

1-2-3 Encuesta, fase 1 [hh/123-1]

1-2-3 Encuesta, fase 2 [hh/123-1]

1-2-3 Encuesta, fase 3 [hh/123-1]

Censo agropecuario [ag/census]

Censo de empresas [en/census]

Censo de población y vivienda [hh/popcen]

Cuestionario de indicadores básicos de bienestar [hh/cwiq]

Encuesta agropecuaria [ag/oth]

Otras encuestas a hogares [hh/oth]

Encuesta de demografía y salud [hh/dhs]

Encuesta de demografía y salud, especial [hh/dhs-sp]

Encuesta de demografía y salud, interina [hh/dhs-int]

Encuesta de demografía y salud, ronda 1 [hh/dhs-1]

Encuesta de demografía y salud, ronda 2 [hh/dhs-2]

Encuesta de demografía y salud, ronda 3 [hh/dhs-3]

Encuesta de demografía y salud, ronda 4 [hh/dhs-4]

Encuesta de empresas [en/oth]

Encuesta de ingresos/gastos/hogares [hh / s]  
Encuesta de mercado laboral [hh/lfs]  
Encuesta de monitoreo socio-económico [hh / sems]  
Encuesta de precios [hh/prc]  
Encuesta de trabajo infantil [hh/clt]  
Encuesta del sector informal [hh / iss]  
Encuesta integrada (excluyendo la Encuesta de Calidad de Vida) [hh / es]

Encuesta mundial de fertilidad [hh / wfs]  
Encuesta mundial de salud [hh / whs]  
Encuesta prioritaria [hh/ps]  
Encuestas de indicadores múltiples por conglomerados - ronda 1 [hh/mics-1]  
Encuestas de indicadores múltiples por conglomerados - ronda 2 [hh/mics-2]  
Encuestas de indicadores múltiples por conglomerados - ronda 3 [hh/mics-3]  
Estudio de medición de las condiciones de vida [hh/lsmc]  
Evaluación de prestación de servicios [hh / spa]

Encuesta mundial de salud [hh / whs]  
Encuesta prioritaria [hh/ps]  
Encuestas de indicadores múltiples por conglomerados - ronda 1 [hh/mics-1]  
Encuestas de indicadores múltiples por conglomerados - ronda 2 [hh/mics-2]  
Encuestas de indicadores múltiples por conglomerados - ronda 3 [hh/mics-3]  
Estudio de medición de las condiciones de vida [hh/lsmc]  
Evaluación de prestación de servicios [hh / spa]

Fuente. Elaboración propia tomada del Nesstar Publisher

**Título traducido:** si el título se encuentra en un idioma diferente al inglés, ingrese la traducción del título al inglés. Si la operación estadística fue desarrollada en un país con más de un idioma oficial, el título puede ser traducido al segundo idioma oficial, en lugar del inglés. Se sugiere que la primera letra debe ir en mayúscula, como por ejemplo: Sample monthly hotels.

**Número de identificación:** corresponde al número o texto único que identifica la operación estadística. Puede utilizar un número de identificación de su elección o la abreviación de la entidad (DANE), abreviación de la dependencia responsable (DIMPE), la de la operación (GEIH) y la fecha de la operación estadística (2010).

Gráfica 10. Identificador de la operación estadística

Identificador de la operación estadística
DANE-DIMPE-GEIH-2007
Descripción del elemento:

Fuente. Elaboración propia tomada del Nesstar Publisher

**Descripción general:** los elementos en este grupo proporcionan un resumen general sobre la operación estadística.

Gráfica 11. Descripción general



Diagrama de flujo que muestra la estructura de una operación estadística, organizada en secciones horizontales:

- Resumen
- Antecedentes de la operación estadística y referentes internacionales
- Objetivos
- Marco de referencia
- Unidades estadísticas
- Clase de datos

Fuente. Elaboración propia tomada del Nesstar Publisher

**Resumen:** esta sección presenta una idea general, pero precisa de los diversos aspectos que componen la operación estadística, es decir, debe presentar de forma clara y resumida una visión general de lo que se trata la operación o investigación estadística. Se trata de realizar un planteamiento claro y ordenado del tema y de la importancia de sus implicaciones, así como de la manera como se han abordado los diferentes elementos que componen la misma.

**Antecedentes de la operación estadística y referentes internacionales:** estos bosquejan el origen y evolución histórica de la operación estadística, señalando sus principales hitos, así como sus modificaciones más relevantes. Lo anterior resulta de vital importancia, para contextualizar a los usuarios, sobre el trabajo adelantado y la experiencia acumulada en la realización de la operación estadística, además permite construir una visión comparativa acerca de los cambios metodológicos que pueden incidir en la obtención y el análisis de sus resultados. En cuanto a los referentes, enuncia los organismos internacionales rectores en la

materia (ONU y sus agencias, EUROSTAT, institutos nacionales de estadística reconocidos a nivel mundial, entre otros) y presenta las principales recomendaciones adoptadas y/o adaptadas por la operación o investigación estadística.

En el caso de nuevas operaciones, resulta importante contextualizar al usuario sobre el origen de la misma, referentes de estudios que aborden temáticas o que cuenten con características similares, y presentar sus conclusiones más relevantes de las pruebas piloto realizadas en el diseño de la operación estadística.

### **Objetivos:**

Establece lo que pretende la operación estadística y la finalidad del tema de investigación.

**Marco de referencia:** el marco de referencia está compuesto por el marco teórico, el marco conceptual, el marco legal y aquellos temas que se considere necesario incluir, con la finalidad de contextualizar la operación estadística de la mejor manera posible.

**a. Marco teórico.** Contiene un resumen de la revisión realizada a la literatura sobre el tema que trata la operación estadística, describe el estado del arte y los aportes que se realizan en el campo temático. Este proceso permite obtener los argumentos necesarios para delimitar el problema de estudio, permite conocer las teorías que ayudan a ubicar la temática, a interpretar mejor los resultados que se obtienen y a buscar la generación de nuevos enfoques en la forma de abordar los problemas.

**b. Marco conceptual.** El marco conceptual describe, explica y establece las relaciones entre los conceptos fundamentales de la operación estadística. Corresponde a los conceptos o términos técnicos básicos utilizados en la misma. Los demás términos que se empleen se relacionan en un glosario, en el campo «Palabra(s) clave». Es importante que los conceptos que se utilicen en la operación estadística sean estandarizados o armonizados conjuntamente con el DANE, teniendo en cuenta los referentes de organismos internacionales como Naciones Unidas, Eurostat o la OCDE, o de otras oficinas de estadística de países destacados. Este ejercicio tiene como fin la comparabilidad, integración e interoperabilidad de la información estadística.

**c. Marco legal.** Describe la normatividad en la que se circunscribe la operación estadística.

**Unidades estadísticas:** en este elemento se describe la unidad de análisis básico (individuos, hogares, establecimientos, empresas, etc.). Estas comprenden: la unidad de observación, la unidad de análisis y la unidad de muestreo. Cada una de las unidades del estudio a documentar debe identificarse con su respectivo título en mayúscula.

EJ: UNIDAD DE OBSERVACIÓN  
Fincas productoras de flores bajo invernadero y a cielo abierto.

UNIDAD DE ANÁLISIS  
Fincas y lotes dedicados al cultivo de flores bajo invernadero y a cielo abierto.

UNIDAD DE MUESTREO  
Fincas productoras de flores bajo invernadero y a cielo abierto.

**Clases de datos:** se describe el tipo de datos (encuesta por muestreo, censo, operación estadística basada en registros administrativos con fines estadísticos, otros) recolectados durante el estudio. La plantilla incluye un vocabulario controlado para este elemento.

Gráfica 12. Clase de datos

Clase de datos
Encuesta por muestreo (ssd)
Censos (cen)
Operación estadística basada en registros administrativos (adm)
Datos agregados (agg)
Series históricas
Datos clínicos (cli)
Datos de transacción o evento (evn)
Datos obtenidos de la observación (obs)
Datos obtenidos por un proceso (pro)
Datos de manejo de tiempo (tbd)

Fuente. Elaboración propia tomada del Nesstar Publisher

**Cobertura temática:** este ítem permite la documentación de los tópicos que fueron cubiertos por la operación estadística.

Gráfica 13. Cobertura temática

Universo de estudio		
Población objetivo		
Contenido temático		
Cuestionarios		
Palabra(s) clave		
Texto	Vocabulario	Vocabulario URI
Clasificación de los tópicos		
Texto	Vocabulario	Vocabulario URI

Fuente. Elaboración propia tomada del Nesstar Publisher

**Universo de estudio y población objetivo:** describe el conjunto de unidades o individuos a las que se refiere la operación estadística o constituyen la colectividad de interés y que satisfacen una definición común.

**Población objetivo:** describe el conjunto de unidades o individuos a los que se refiere la operación estadística. Estas unidades son delimitadas en términos de contenido, espacio y tiempo. Esta descripción incluye un listado de todos los grupos excluidos de la operación, y de ser relevante, el motivo por el cual se excluyen. Al igual que los anteriores elementos los títulos deben ir en mayúscula.

**Contenido temático:** menciona las variables (o bloques de variables) más importantes de la operación estadística. Para el caso de aquellas investigaciones que presentan un gran número de variables, es necesario clasificarlas en los principales temas de la misma. Para diferenciar cada uno de los temas, los títulos deben quedar en mayúscula.

**Cuestionarios:** identifica por nombre y tipo de cuestionario (no el contenido del cuestionario) y los módulos principales. Se deben usar uno de los siguientes tres términos para describir el tipo de instrumento de recolección de datos utilizado:

- Estructurado: indica un instrumento en el que a todos los encuestados se les aplicó las mismas preguntas/pruebas, posiblemente con respuestas precodificadas. Si una porción pequeña del cuestionario incluye preguntas abiertas, provea los comentarios apropiados.
- Semiestructurado: indica que el documento de investigación contiene, principalmente, preguntas abiertas.
- No estructurado: indica que se condujeron entrevistas exhaustivas.

Los títulos se deben ir en mayúscula. En el documento metodológico se puede encontrar como Marco conceptual y/o Base conceptual pero el título debe quedar como CONTENIDO TEMÁTICO, lo mismo para cuestionarios si aparece en el documento metodológico como Diseño de Instrumentos, cuestionarios o formularios el título debe quedar como CUESTIONARIOS.

**Palabra(s) clave(s):** el objetivo es ayudar a individuos y organizaciones que deseen buscar y clasificar proyectos archivados. Aunque en la plantilla de DDI no existe ningún vocabulario controlado para este elemento, las organizaciones podrán crear una lista unificada de palabras clave para asegurar consistencia entre los proyectos.

Gráfica 14. Palabras claves

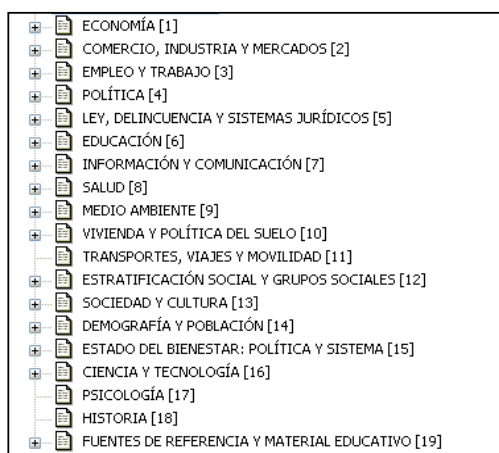
Palabra(s) clave		
Texto	Vocabulario	Vocabulario URI

Fuente. Elaboración propia tomada del Nesstar Publisher



**Clasificación de los tópicos:** esta casilla muestra un listado amplio de tópicos cubiertos por los datos. El listado de términos suministra a los usuarios un listado de valores predeterminados a escoger. El atributo «vocabulario URI» especifica la ubicación del vocabulario controlado.

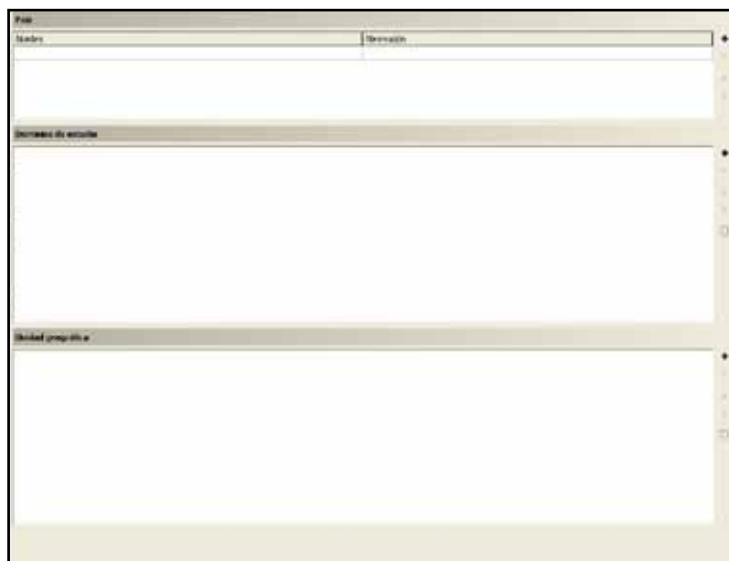
Gráfica 15. Clasificación de los tópicos



Fuente. Elaboración propia tomada del Nesstar Publisher

**Cobertura geográfica:** los elementos en este grupo pueden ser utilizados para proporcionar una descripción del área geográfica y la población cubierta por esta operación estadística.

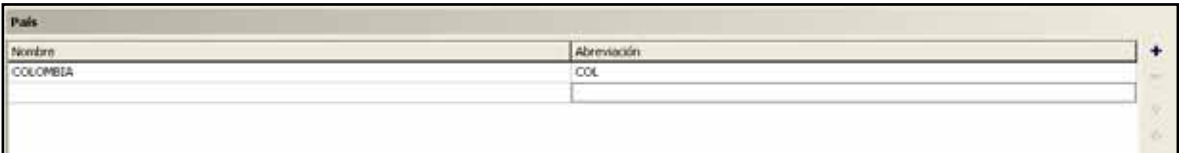
Gráfica 16. Cobertura geográfica



Fuente. Elaboración propia tomada del Nesstar Publisher

**País:** todos los países incluidos en la operación estadística deben estar listados. La plantilla DDI utiliza un vocabulario controlado para este elemento. Se pone el nombre del país en mayúscula y su respectiva abreviación en código ISO que corresponde a las tres primeras letras.

Gráfica 17. País



Nombre	Abreviación
COLOMBIA	COL

Fuente. Elaboración propia tomada del Nesstar Publisher

**Dominios de estudio:** consiste en un subgrupo de la población objetivo sobre el que se presentan resultados y pueden darse en un área geográfica como una región o un centro poblado. También podría incluir una categoría de la población específica, como un gran grupo nacional o étnico.

En las encuestas por muestreo, el número de dominios tiene una influencia importante en el tamaño y la distribución de la muestra. Normalmente, las estadísticas se presentan para los distintos subgrupos de la población, llamados dominios de estudio.

Los dominios de estudio elegidos pueden coincidir con el estrato de muestreo estratificado o grupos que se encuentran en ellos. Estos dominios pueden ser geográficos o no geográficos. A menudo, estos subgrupos están relacionados con una clasificación, por ejemplo, las unidades territoriales, la actividad económica, etc., (Eurostat, “Evaluación de la calidad de las estadísticas: Glosario”, Grupo de Trabajo, Luxemburgo, octubre de 2003).

**Unidad geográfica:** se usará este elemento para proporcionar información acerca de las áreas que fueron cubiertas y para mencionar las razones por las que la encuesta no cubrió el área geográfica total del país. Especificar el nivel más pequeño de agregación cubierto por los datos.

**Productores y auspiciadores:** este grupo contiene elementos que pueden ser utilizados para reconocer la labor de personas y organizaciones responsables del diseño, implementación y financiamiento de la operación estadística. Este grupo cuenta con cuatro elementos:

Gráfica 18. Productores y auspiciadores

Investigador principal			
Nombre	Afiliación		
Colaboradores			
Nombre	Abreviación	Afiliación	Rol
Financiamiento			
Agencia/Organización	Abreviación	Número de Donación	Rol
Reconocimientos			
Nombre	Afiliación		Rol

Fuente. Elaboración propia tomada del Nesstar Publisher

**Investigador principal:** es el responsable del diseño y conducción de la operación estadística. Cabe recordar que los estudios pueden tener más de un investigador principal. También se puede poner el nombre del área productora de la investigación.

Ejemplo: Dirección de Metodología y Producción Estadística (DIMPE)

**Colaboradores:** personas u organizaciones involucradas en las distintas etapas de la operación estadística (diseño, levantamiento de campo, procesamiento y análisis). Se debe diligenciar de la siguiente manera:

Nombre: nombre de la Dirección - nombre de la operación estadística (Ubicar primeras letras en mayúscula).

Abreviación: en mayúscula la abreviación de la Dirección – operación estadística.

Afiliación: entidad con las primeras en mayúscula y sigla.

Rol: equipo técnico en mayúscula las primeras letras.

(Ver tabla 1).

Tabla 1. Colaboradores

Nombre	Abreviación	Afiliación	Rol
Dirección de Metodologías y Producción Estadística - Gran Encuesta Integrada De Hogares	DIMPE - GEIH	Departamento Administrativo Nacional de Estadística (DANE)	Equipo Técnico

Fuente. Elaboración propia

**Financiamiento:** las organizaciones responsables de financiar la operación estadística deberían ser listadas en este. Si es posible, reporte el número de donación o contrato.

Agencia: nombre de la entidad.

Abreviación: sigla de la entidad.

Rol: ejecutor cuando se es el productor de la operación con recursos propios (Ver tabla 2).

Tabla 2. Financiamiento

Agencia	Abreviación	Número de Donación	Rol
Departamento Administrativo Nacional de Estadística (DANE)	DANE		Ejecutor

Fuente. Elaboración propia

**Reconocimientos:** personas y/o instituciones que contribuyeron a la producción de la operación estadística: asistencia técnica, estudios de referencia, etc.

**Muestreo:** en este ítem se debe documentar el diseño y definición del tamaño de la muestra, incluyendo marco muestral, tipo de muestreo y tamaño final de la muestra, como también, pérdida de muestra, procedimiento de estimación y cálculo de precisión de los resultados. Este ítem solo aplica para encuestas por muestreo.

Gráfica 19. Muestreo



Diagrama de flujo de cuatro etapas de muestreo:

- Diseño y definición del tamaño de la muestra** (Etapa 1)
- Pérdida de muestra** (Etapa 2)
- Procedimiento de estimación** (Etapa 3)
- Cálculo de precisión de los resultados** (Etapa 4)

El diagrama muestra una secuencia de cuatro rectángulos blancos con bordes grises, cada uno con un título en mayúscula. Los rectángulos están conectados por líneas grises que indican el flujo de la información. En la parte superior derecha del primer rectángulo hay un signo '+'. En la parte superior derecha de los rectángulos 2, 3 y 4 hay un signo '3'.

Fuente. Elaboración propia tomada del Nesstar Publisher

**Diseño y definición del tamaño de la muestra:** este elemento proporciona información sobre el marco muestral y los métodos y procedimientos utilizados para seleccionar a los encuestados. El tamaño de muestra deseado también debe mencionarse. Los títulos de cada uno de los temas deben quedar en mayúscula.

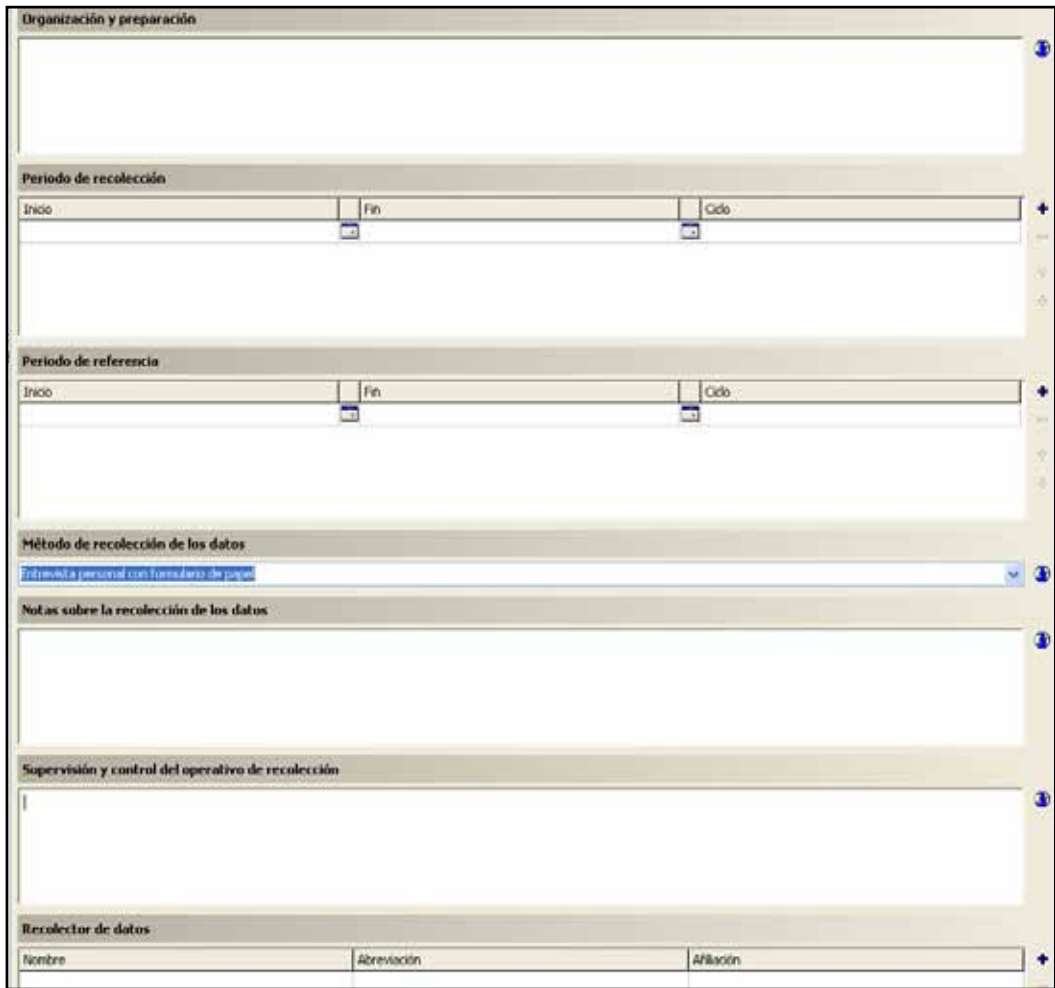
**Pérdida de muestra:** este elemento se usará para describir la correspondencia entre las unidades que fueron encuestadas exitosamente y la muestra planeada. Cualquier desviación importante debe ser mencionada aquí.

**Procedimiento de estimación:** explica los estimadores que se utilizan para obtener los indicadores de la operación estadística. Define y justifica la metodología elegida y sus componentes. Indica el cálculo de los factores de expansión y los elementos que determinan la expansión. Cada uno de los títulos que se utilicen en la documentación de este campo debe ir en mayúscula.

**Cálculo de precisión de los resultados:** describe la metodología utilizada para estimar los errores de muestreo y su presentación, con el fin de determinar el nivel de confianza. Reseña el diseño del método de estimación de la varianza. Presenta las fórmulas para calcular el error estándar y/o el coeficiente de variación de los estimadores. Se deben poner en mayúscula los subtítulos que se utilicen en la documentación de este campo.

**Recolección de datos:** una operación estadística correctamente documentada estará acompañada por un informe completo sobre las actividades de recolección de datos y una copia de todos los cuestionarios utilizados durante su ejecución. Este grupo de elementos tiene la intención de proveer de un breve resumen de las actividades de recolección de los datos.

Gráfica 20. Recolección de datos



El formulario de recolección de datos está dividido en varias secciones:

- Organización y preparación:** Sección vacía para notas.
- Periodo de recolección:** Campos para Inicio, Fin y Cód.
- Periodo de referencia:** Campos para Inicio, Fin y Cód.
- Método de recolección de los datos:** Menú desplegable con la opción "Entrevista personal con formularios de papel" seleccionada.
- Notas sobre la recolección de los datos:** Área de texto para notas.
- Supervisión y control del operativo de recolección:** Área de texto para notas.
- Recolector de datos:** Campos para Nombre, Abreviación y Afiliación.

Fuente. Elaboración propia tomada del Nesstar Publisher

**Organización y preparación:** se deben incorporar los elementos relacionados con los procesos de sensibilización, capacitación y el esquema operativo. Los subtítulos deben ir en mayúscula.

**Periodo de recolección:** ingrese las fechas de inicio y fin de la recolección de datos indicando el ciclo del mismo (mensual, trimestral, semanal, etc.). Si no se tiene el mes ni el día, se puede ingresar solo el año. Las fechas deben ser ingresadas en formato ISO, ej. AAAA-MM-DD (o AAAA-MM, o AAAA). Es importante tener en cuenta que esta fecha debe ser posterior a la reportada en el periodo de referencia.

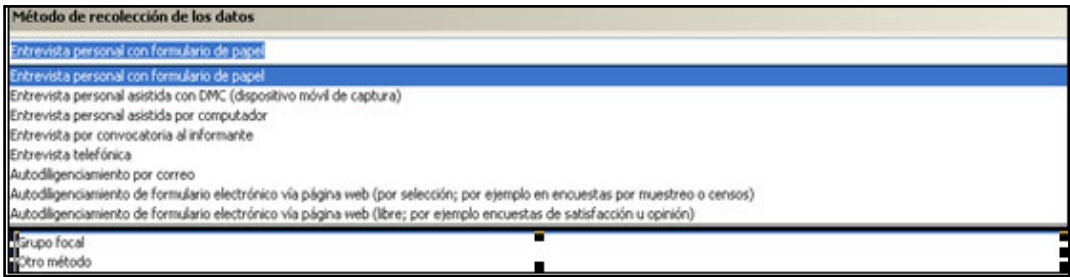
Tabla 3. Periodo de recolección

Inicio	Término	Ciclo
2009 - 11 - 01	2009 - 12 - 30	Bimestral

Fuente. Elaboración propia

**Método de recolección:** seleccione el modo principal de recolección de datos del listado proporcionado en este elemento.

Gráfica 21. Método de recolección



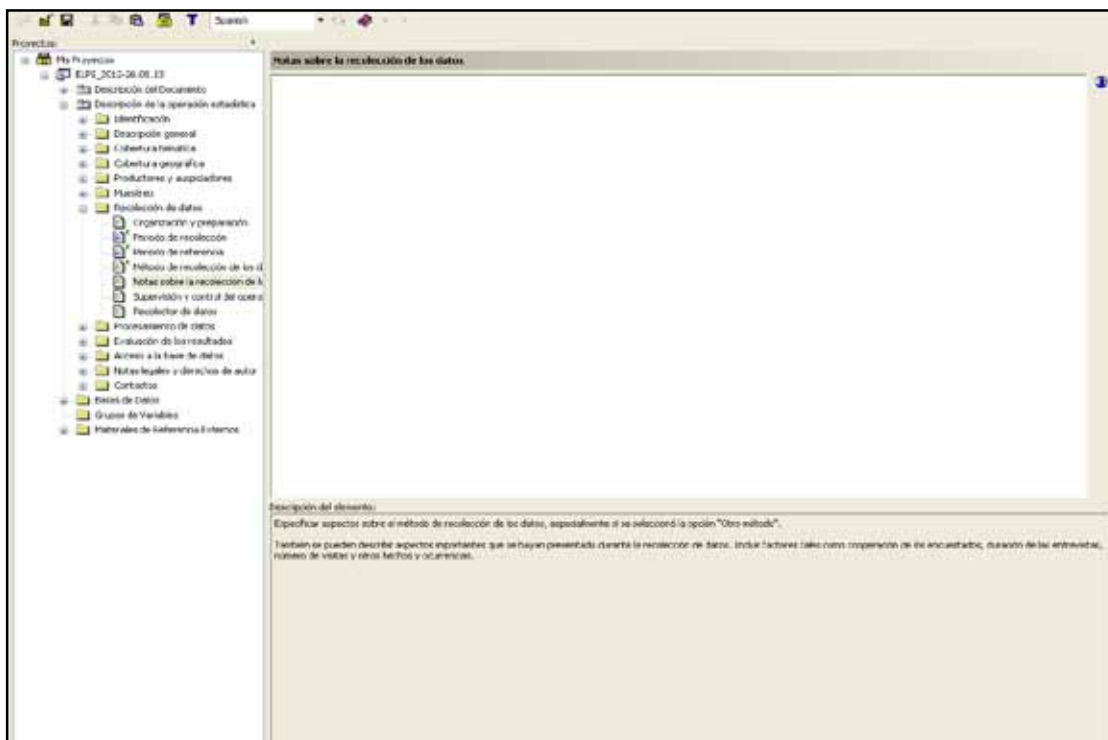
Método de recolección de los datos

- Entrevista personal con formulario de papel
- Entrevista personal con formulario de papel
- Entrevista personal asistida con DMC (dispositivo móvil de captura)
- Entrevista personal asistida por computador
- Entrevista por convocatoria al informante
- Entrevista telefónica
- Autodiligenciamiento por correo
- Autodiligenciamiento de formulario electrónico vía página web (por selección; por ejemplo en encuestas por muestreo o censos)
- Autodiligenciamiento de formulario electrónico vía página web (libre; por ejemplo encuestas de satisfacción u opinión)
- Grupo focal
- Otro método

Fuente. Elaboración propia tomada del Nesstar Publisher

**Notas sobre la recolección de datos:** se deben especificar aspectos sobre el método de recolección de los datos, especialmente si se seleccionó la opción «Otro método». También se pueden describir aspectos importantes que se hayan presentado durante la recolección de datos. Incluir factores tales como cooperación de los encuestados, duración de las entrevistas, número de visitas y otros hechos y ocurrencias.

Gráfica 22. Notas sobre la recolección de los datos



Notas sobre la recolección de los datos

Descripción del elemento:  
Especificar aspectos sobre el método de recolección de los datos, especialmente si se seleccionó la opción "Otro estado".  
También se pueden describir aspectos importantes que se hayan presentado durante la recolección de datos. Incluir factores tales como cooperación de los encuestados, duración de las entrevistas, número de visitas u otros hechos u ocurridos.

Fuente. Elaboración propia tomada del Nesstar Publisher

**Supervisión y control del operativo de recolección:** presenta los controles que se aplican en terreno al personal a cargo del supervisor o coordinador de campo, tales como: recorridos, cobertura de las áreas geográficas y de unidades asignadas, equipos y dispositivos, seguridad de los cuestionarios diligenciados. También incluye los controles para reducir el sesgo, en situaciones como: rechazos totales o parciales; no respuesta; falta de cobertura; revisitas; errores en el marco y novedades; no existencia de la fuente; error de dirección; cambio de dirección; diligenciamiento de encuestas de cobertura y verificación de la veracidad de la información recolectada con las variables más relevantes, para evaluación de la calidad. También incluirá un resumen de los procedimientos seguidos para minimizar errores, reducir pérdida de información y prevenir falsificación. Estos procedimientos pueden involucrar revisiones de los supervisores de los cuestionarios diligenciados en el campo, re-entrevistas a los encuestados y otras actividades similares.

**Periodo de referencia:** indica y justifica el intervalo de tiempo (años, meses, semanas, días) a los que se refiere la información, indicando el ciclo del mismo (mensual, trimestral, semanal, etc.). Si solo se tiene el año se puede ingresar sin mes ni día.



Se ingresan las fechas en formato ISO, ej. AAAA-MM-DD (o AAAA-MM, o AAAA). Es importante tener en cuenta que esta fecha debe ser anterior a la reportada en el periodo de recolección.

Tabla 4. Periodo de referencia

Inicio	Término	Ciclo
2009 - 07 - 01	2009 - 09 - 30	Trimestral

Fuente. Elaboración propia

**Recolector de datos:** proporciona un listado de las personas y las organizaciones responsables de administrar los cuestionarios y recolectar los datos. Esto se refiere a entidades que recolectan los datos, no a las que producen la documentación. Las primeras letras del nombre de la entidad deben ir en mayúscula, la abreviación de la entidad en mayúscula y la afiliación. La afiliación, corresponde a Gobierno Nacional para entidades públicas del orden nacional, para el caso de las secretarías departamentales su afiliación será departamental y para el caso de los gremios la afiliación será gremial. Ver tabla 5.

Tabla 5. Recolección de datos

Nombre	Abreviación	Afiliación
Departamento Administrativo Nacional de Estadística	DANE	Gobierno Nacional

Fuente. Elaboración propia

**Procesamiento de datos:** la validación de los datos es el único elemento en este grupo. En este apartado se ingresa un resumen general de los procedimientos utilizados para identificar y corregir errores en los datos. Dentro de los temas que se incluyen están:

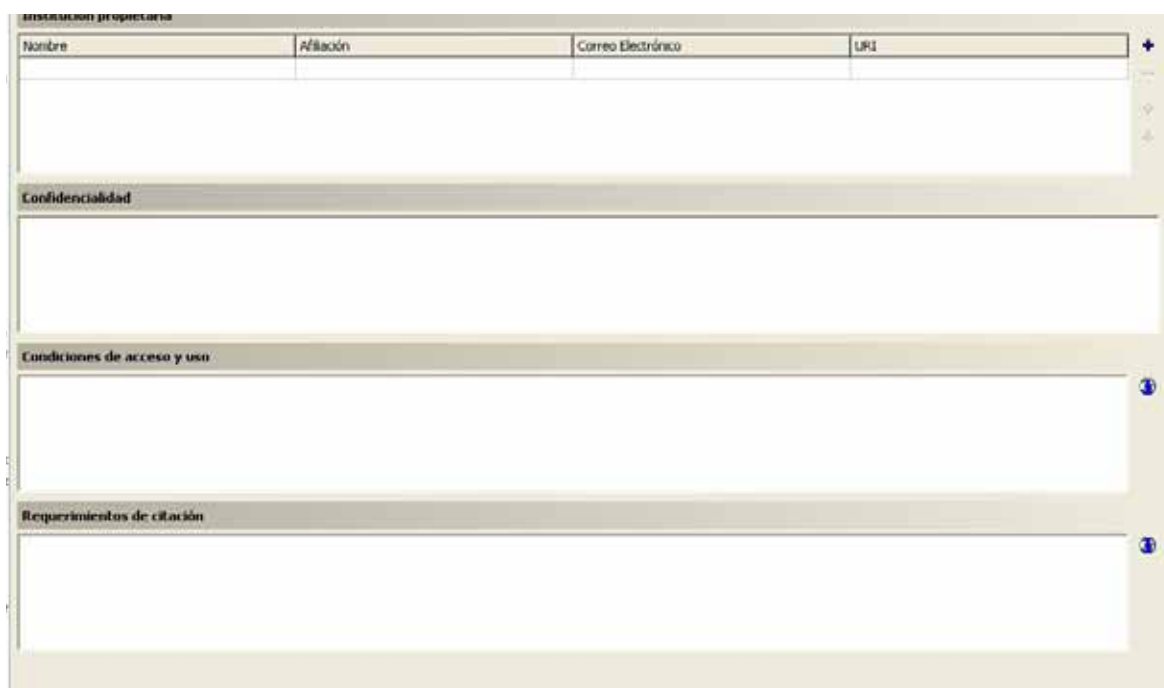
- La consolidación de archivos.
- Las normas de validación y consistencia.
- La verificación de la consistencia interna de los datos y ajustes.
- La imputación y/o ajustes de cobertura.

**Evaluación de los resultados:** describe los estimadores de precisión de los datos de la operación estadística (error de muestreo, coeficientes de variación, intervalos de confianza, etc.).

**Otras formas de validación de los datos:** se refiere a otras observaciones relacionadas con la validación de los datos. Por ejemplo, la variancia de las respuestas; las pruebas del sesgo de la muestra; el sesgo del encuestador o la respuesta; los intervalos de confianza, etc.

**Acceso a las bases de datos:** los elementos en este grupo proporcionan un resumen de las condiciones bajo las cuales se puede acceder a un estudio de datos.

Gráfica 23. Acceso a las bases de datos



Nombre	Afiliación	Correo Electrónico	URI

**Confidencialidad**

**Condiciones de acceso y uso**

**Requerimientos de citación**

Fuente. Elaboración propia tomada del Nesstar Publisher

**Institución propietaria:** debe contener los nombres de las personas u organizaciones responsables de otorgar acceso a los datos y documentación.

Tabla 6. Institución propietaria

Nombre	Afiliación	Correo Electrónico	URL
Departamento Administrativo Nacional de Estadística	Gobierno Nacional	dane@dane.gov.co	www.dane.gov.co

Fuente. Elaboración propia

**Confidencialidad:** describe las normas que establecen el anonimato del informante y los compromisos de la institución para garantizar la confidencialidad de los resultados.

**Condiciones de acceso y uso:** proporciona una descripción de los términos bajo los cuales se permite a los usuarios el acceso a los datos de la operación estadística. Por ejemplo, algunas bases de datos pueden ser obtenidas gratuitamente en Internet, mientras que a otras únicamente se puede acceder desde computadoras designadas en instalaciones especiales.

**Requerimientos de citación:** presenta la acreditación de la fuente de datos conforme a las normas de procedimientos de citación del país respectivo.

**Notas legales:** proveer información relativa a la responsabilidad de los usuarios de la documentación y las bases de datos.

**Derechos de autor:** se refiere a la declaración de derechos de autor de los resultados, con el fin de respetar la autoría y la propiedad intelectual, otorgando protección tanto a las fuentes de información como a la entidad generadora de los indicadores.

**Contactos:** este grupo puede ser utilizado para proporcionar a los usuarios información acerca de quién puede responder preguntas adicionales sobre la operación estadística.

Gráfica 24. Contactos



Personas a contactar			
Nombre	Afiliación	Correo Electrónico	URL

Fuente. Elaboración propia tomada del Nesstar Publisher

**Personas a contactar:** se debe ingresar el nombre y la forma de contactar de las personas o agencias responsables de la operación estadística. Estas personas son usualmente los investigadores principales, pero se podría incluir a otras personas que puedan responder preguntas acerca del diseño la operación y la recolección de datos.

**6.3. Base de datos**

Esta sección contiene todos los archivos de datos asociados con la operación estadística, además de una descripción detallada de cada variable en los archivos de datos. Los elementos en esta sección son obtenidos de las especificaciones DDI. Cada archivo de datos tiene cuatro grupos de elementos: descripción del archivo, variables clave y relaciones, variables y entrada de datos.

**Descripción del archivo:** los elementos en este grupo son utilizados para proporcionar información básica acerca de cada archivo. La Plantilla DDI cuenta con los siguientes elementos en este grupo:

Gráfica 25. Descripción del archivo

Contenido	
Productor	
Versión	
Datos perdidos	
Notas	

Fuente. Elaboración propia tomada del Nesstar Publisher

**Contenido:** este elemento debe referirse solamente al archivo de datos seleccionado y no a la operación estadística. Proporciona una descripción de las bases de datos incluyendo cobertura temática, características especiales de su contenido y principales variables.

**Productor:** se ingresa el nombre de la persona u organización responsable de la creación de la base de datos. Se debe tener presente que el productor del archivo de datos no es necesariamente la agencia que recolectó o procesó los datos.

Ejemplo: Departamento Administrativo Nacional de Estadística (DANE)

**Versión:** se ingresa una descripción de cada versión del archivo de datos en este elemento. La información debe incluir fecha, origen y tipo de datos (Validados, Parcialmente validados, originales, etc.). Idealmente, un número de versión debe ser asignado a cada archivo de datos. Se recomienda que una numeración formal sea utilizada para identificar diferentes publicaciones de la base de datos. Se debe poner la palabra «versión» y el año de la operación estadística, así: Versión 2005.

**Datos perdidos:** se emplea este elemento para describir las causas por la que existen datos perdidos en el archivo de datos. Los datos perdidos particulares son definidos al documentar cada variable.

*Notas:* se puede ingresar algún comentario adicional acerca del archivo de datos.

**Variables claves y relaciones:** la mayoría de las operaciones estadísticas incluyen múltiples archivos de datos relacionados. Este grupo de elementos está diseñado con el propósito de definir las relaciones entre estos archivos. Son usados dos tipos de archivos para definir las relaciones entre los archivos de datos: variables clave base y externas.

- **Variables clave base:** son variables que identifican individualmente, cada observación en el archivo de datos. Por ejemplo, una variable clave base en el archivo de hogares sería la variable que contiene un número de identificación único para cada hogar. Si el archivo de datos no contiene esta variable, una nueva puede ser creada mediante una combinación única de variables; por ejemplo, la combinación de variables como región, área y número de identificación del hogar en el área.

- **Variable clave externa:** se refiere a aquellos casos en los que las variables que pueden ser utilizadas para relacionar bases de datos no incluyen una variable con un número de identificación único para cada observación. Estas variables permiten que una fusión de archivos no contenga valores duplicados.

Es importante crear variables clave y definir las relaciones entre bases de datos en un proyecto correctamente. Pues las «variables clave base» serán utilizadas por los analistas de datos para fusionar los archivos en STATA, SPSS, SAS o cualquier otro programa estadístico.

**Variables:** este grupo de elementos es utilizado para ingresar metadatos relacionados con cada variable.

- **Listado de variables:** debe contener la siguiente información acerca de todas las variables en las bases de datos:

Gráfica 26. Listado de variables

Variables				
Variables:				
Número	Nombre	Etiqueta	Áncho	Decimales

Fuente. Elaboración propia tomada del Nesstar Publisher

- El nombre es el medio principal para referirse a una variable. No debe tener más que 8 caracteres, no debe empezar con un número y no debe contener espacios en blanco.
- La etiqueta es la descripción de la variable. Esta descripción debe ser breve, pero descriptiva.
- El ancho es el máximo número de caracteres que pueden ser incluidos en una variable.
- Los decimales corresponde al número de lugares utilizados por decimales en cada variable.

**Descripción de la variable:** esta opción cuenta con campos para definir las etiquetas de los valores o categorías de una variable y para ingresar otra información, como datos perdidos o nivel de medición. Las categorías permiten definir etiquetas de los valores de una variable.

**Documentación:** proporciona información acerca de la base de datos y permite crear y guardar resúmenes estadísticos como metadatos, lo que facilita el ingreso de información más detallada sobre las variables que lo requieran.

**Estadísticas:** el Nesstar Publisher calcula y guarda varias medidas estadísticas comunes como: número de casos válidos ponderados y sin ponderar; valores mínimos y máximos; y media y desviación estándar ponderada y sin ponderar. Estas estadísticas son guardadas como metadatos y esto hace posible que se reporten frecuencias y otras medidas de resumen sin tener que publicar los microdatos.

**Ponderadores:** al documentar las operaciones estadísticas, se puede especificar las variables como ponderadores y aplicar esta ponderación a variables individuales cuando sea apropiado. Después se podrá calcular y mostrar estadísticas ponderadas en el resumen de estadísticas.

**Documentación:** la documentación de la variable incluida en los estándares estadísticos de los archivos de datos está, a menudo, limitada a la etiqueta de la variable y las etiquetas de los valores. El Nesstar Publisher proporciona más elementos para que la documentación sea más completa, de la siguiente manera:

Gráfica 27. Documentación

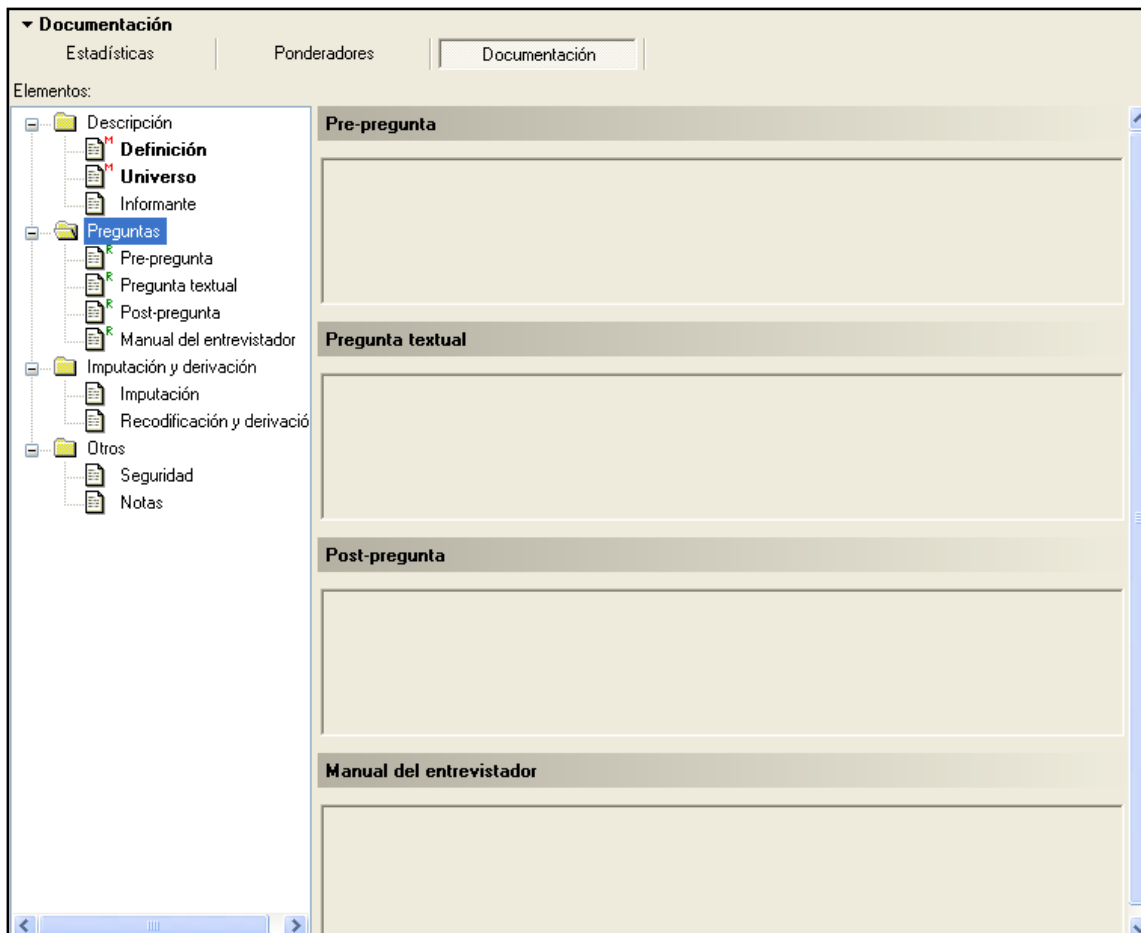


Fuente. Elaboración propia tomada del Nesstar Publisher

## Descripción

- El elemento **definición** permite que la variable sea descrita en mayor detalle que el proporcionado por la etiqueta variable.
- El elemento **universo** permite que los usuarios especifiquen la población exacta a la que la variable es aplicable.
- El elemento **encuestado idóneo** debe documentar información respecto a la persona/ entidad que proporciona la información de la variable. En encuestas de hogares, la fuente podrá ser el jefe del hogar o un miembro del hogar. Pero también puede ser el entrevistador a través de observaciones visuales, el gerente o representante legal de una empresa, etc.

Gráfica 28. Preguntas



The screenshot displays the 'Preguntas' (Questions) section of a software interface. On the left, a tree view under 'Documentación' shows the following structure:

- ▼ Documentación
  - Estadísticas
  - Ponderadores
  - Documentación
    - Elementos:
      - Descripción
        - Definición
        - Universo
        - Informante
      - Preguntas
        - Pre-pregunta
        - Pregunta textual
        - Post-pregunta
        - Manual del entrevistador
      - Imputación y derivación
        - Imputación
        - Recodificación y derivación
      - Otros
        - Seguridad
        - Notas

The main area on the right contains four sections, each with a large text input field:

- Pre-pregunta**
- Pregunta textual**
- Post-pregunta**
- Manual del entrevistador**

Fuente. Elaboración propia tomada del Nesstar Publisher

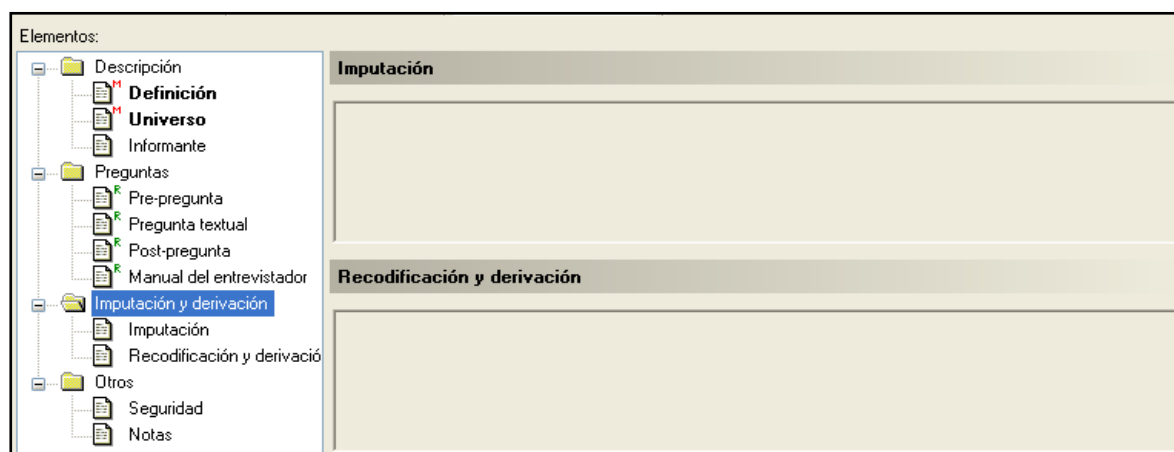


## Preguntas:

- El elemento «Texto» de la **Pre-pregunta** puede ser utilizado para describir la pregunta inmediatamente anterior, según las condiciones del formulario.
- Ingrese la redacción exacta de la pregunta en el elemento de **Pregunta textual**. Este elemento es el más importante en la sección de «Descripción» de la documentación de la variable.
- En el elemento de «Texto» de la **Pos-pregunta** se ingresa algún texto que sigue a la pregunta en el cuestionario y que sirve como guía al entrevistador.
- El elemento de **Manual del entrevistador** debe contener una guía para el entrevistador. Esta información será, típicamente, copiada del manual del entrevistador.

## Imputación y derivación:

Gráfica 29. Imputación y derivación



Fuente. Elaboración propia tomada del Nesstar Publisher

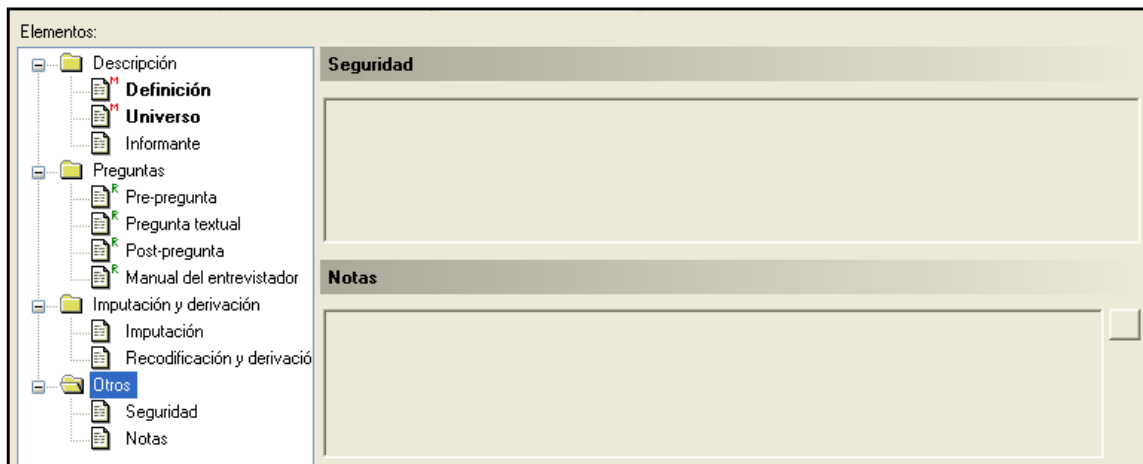
Muchos archivos de datos incluirán variables derivadas o generadas, además de las variables con información recolectada. Asimismo, algunas variables pueden incluir valores imputados. Documentar tales imputaciones es crucial para generar confianza y asegurar que los usuarios pueden reproducir la construcción de datos.

Si los datos perdidos han sido reemplazados con estimaciones, entonces el proceso utilizado para realizar estas estimaciones debería ser descrito con tanto detalle como fuese posible en el elemento de imputación. Este elemento puede incluir una referencia a un documento técnico más detallado.

Las variables también pueden ser obtenidas recodificando o combinando otras variables. En tal situación, el elemento de «Recodificación» y «Derivación» debe contener una descripción clara y completa de todas las acciones realizadas al preparar la variable.

### Otros:

Gráfica 30. Otros



Fuente. Elaboración propia tomada del Nesstar Publisher

- Use el elemento **Seguridad** para describir el nivel del acceso apropiado para una variable.
- El elemento **Notas** puede ser utilizado para indicar alguna otra información acerca de la variable no mencionada en ninguna otra parte.

**Grupo de variables:** los archivos de datos pueden incluir cientos de variables. El Nesstar Publisher proporciona una herramienta para organizar variables en grupos. La agrupación de variables permite que los usuarios naveguen a través de los listados de variables rápidamente y ayuda a controlar el análisis proporcionando una indicación de qué artículos en una base de datos están conceptualmente conectados. Una variable puede pertenecer a más de un grupo y un grupo de variables puede incluir variables de más de un archivo de datos.

La definición de grupo de variables es un proceso de dos pasos.

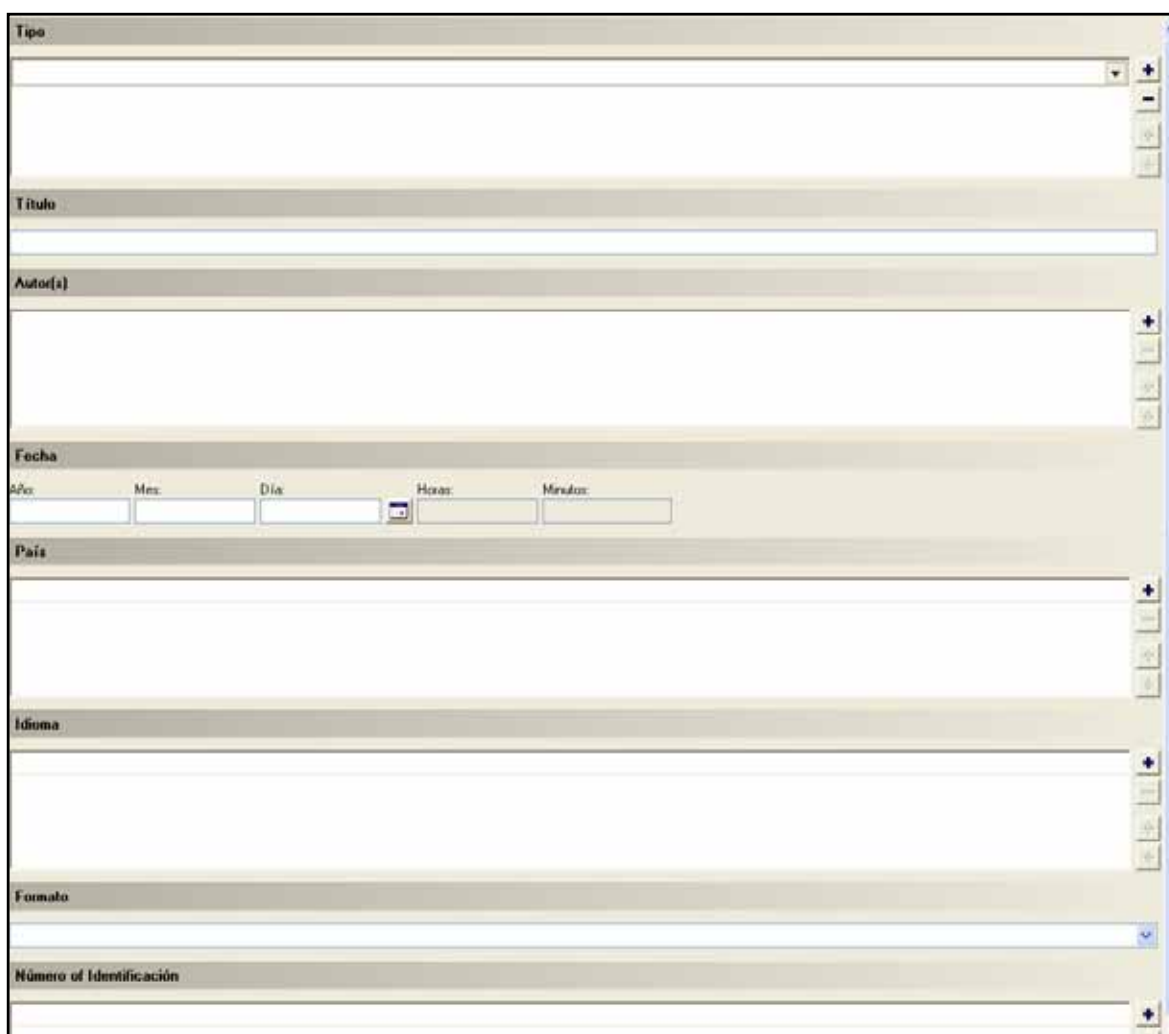
Paso 1: el grupo es creado.

Paso 2: las variables son agregadas al grupo.

## 6.4. Materiales de referencia externos

Esta sección se utiliza para proporcionar un listado y una descripción de materiales como: documentos (manuales, cuestionarios, informes técnicos y analíticos), programas de computadora (entrada de datos, edición, tabulación, análisis), y fotos y mapas relacionados de la operación estadística. En esta se usa la especificación del DCMI y, a diferencia de los archivos de datos, los materiales de referencia externos no son guardados en el archivo de Nesstar Publisher. Solamente los metadatos que describen estos recursos son guardados con el proyecto.

Gráfica 31. Materiales de referencia externos



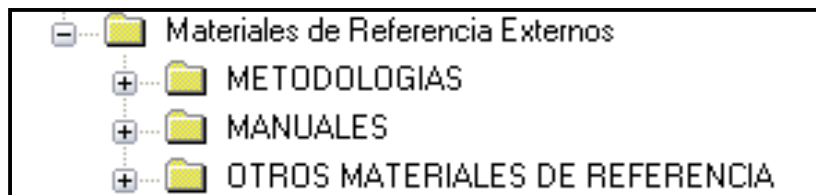
El formulario muestra los siguientes campos:

- Tipo:** Campo de selección con una lista desplegable y botones de expansión (+/-).
- Título:** Campo de texto para el título principal.
- Autor(a):** Campo de texto para el autor, con botones de expansión (+/-).
- Fecha:** Campos separados para Año, Mes, Día, Horas y Minutos.
- País:** Campo de texto para el país, con botones de expansión (+/-).
- Idioma:** Campo de texto para el idioma, con botones de expansión (+/-).
- Formato:** Campo de texto para el formato del archivo, con botones de expansión (+/-).
- Número de Identificación:** Campo de texto para un identificador único, con botones de expansión (+/-).

Fuente. Elaboración propia tomada del Nesstar Publisher

El material de referencia se puede clasificar en diferentes carpetas. Estas se deben nombrar de acuerdo con los tipos de documentos que van a contener, así como se muestra en la siguiente gráfica:

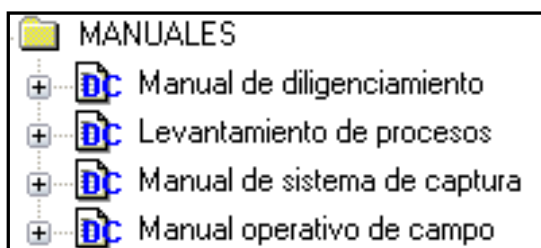
Gráfica 32. Ejemplo de presentación de material de referencia externo



Fuente. Elaboración propia tomada del Nesstar Publisher

De acuerdo con lo anterior, el nombre de los documentos que contenga cada una de las carpetas, debe iniciar en mayúscula y el mismo debe ser representativo, como se muestra a continuación.

Gráfica 33. Ejemplo de presentación de los nombres de los documentos



Fuente. Elaboración propia tomada del Nesstar Publisher

**Identificación:** este grupo consta de ocho elementos, que se indican a continuación:

- 1. Tipo:** este elemento se usa para indicar la clase de material de referencia que está siendo documentado.

Gráfica 34. Tipo de material de referencia

Tipo
Document, Administrative [doc/adm]
Document, Analytical [doc/ani]
Document, Other [doc/oth]
Document, Questionnaire [doc/qst]
Document, Reference [doc/ref]
Document, Report [doc/rep]
Document, Technical [doc/tec]
Audio [aud]
Map [map]
Photo [pic]
Program [prg]
Table [tbl]
Video [vid]
Web Site [web]
Database [dat]

Fuente. Elaboración propia tomada del Nesstar Publisher

- 2. Título y subtítulo:** en estos elementos se debe ingresar un nombre formal y otro de manera opcional que es secundario al material de referencia.
- 3. Autor:** en este elemento se escribe el nombre de las personas individuales o las organizaciones responsables de crear el material de referencia. Se debe usar nombres completos y se recomienda escribir las siglas o abreviaturas dentro de paréntesis después del nombre de la organización.
- 4. Fecha (opcional):** en este elemento se ingresa la fecha completa o parcial en que el material de referencia fue creado o fue modificado por última vez.
- 5. País:** en este elemento se listan todos los países dentro del alcance del material de referencia en el elemento.
- 6. Idioma:** en este elemento se listan todos los idiomas en que aparecen los materiales de referencia.
- 7. Formato:** se debe seleccionar un objeto del listado para identificar el formato del archivo del material de referencia.
- 8. Número de identificación:** si existe se debe presentar el identificador del material de referencia.

## Contribuidores y derechos legales:

**Contribuidores:** en este elemento es necesario citar a las personas u organizaciones que han apoyado o contribuido al desarrollo del material de referencia.

**Derechos:** se utiliza para proporcionar información acerca de los derechos legales de los materiales de referencia.

## Contenido:

**Descripción:** en este elemento se debe describir brevemente el contenido del material de referencia.

**Resumen:** proporciona información resumida sobre cada uno de los principales aspectos del material de referencia.

**Tabla de contenido:** este elemento es usado para listar todas las secciones del informe, cuestionario y otro documento.

## BIBLIOGRAFÍA

Departamento Administrativo Nacional de Estadística (DANE). (2012a). Guía para documentar y codificar documentación técnica. Sistema Integrado de Gestión Institucional. Bogotá: DANE.

—. (2012b). Guía para la elaboración de protocolos para la actividad estadística. DIRPEN. Versión 2.0. Bogotá: DANE.

—. (2010c). Instructivo para la elaboración de guías. Sistema Integrado de Gestión Institucional. Bogotá: DANE.

—. (2012d). Instructivo para la elaboración de guías. Sistema Integrado de Gestión Institucional. Bogotá: DANE.

—. (2012e). Resolución 1503 de 2011. Bogotá: DANE.

Eurostat. Evaluación de la calidad de las estadísticas: Glosario, Grupo de Trabajo, Luxemburgo, octubre de 2003.

Organización para la Cooperación y el Desarrollo Económico (OCDE) (s.f.). Glosario de términos estadísticos. Recuperado en 2012, de:  
<http://stats.oecd.org/glossary/detail.asp?ID=532>.