

**DIRECCIÓN DE REGULACIÓN, PLANEACIÓN,
ESTANDARIZACIÓN Y NORMALIZACIÓN
-DIRPEN-**

COORDINACIÓN DE REGULACIÓN ESTADÍSTICA

**GUÍA PARA LA IMPLEMENTACIÓN DE LOS ESTÁNDARES
ESTADÍSTICOS PARA LA DOCUMENTACIÓN DE
METADATOS Y MICRODATOS (DATA DOCUMENTATION
INITIATIVE (DDI) Y DUBLIN CORE (DC))**

Junio – 2020

Versión 2.0

**DEPARTAMENTO ADMINISTRATIVO
NACIONAL DE ESTADÍSTICA
(DANE)**

JUAN DANIEL OVIEDO ARANGO
Director

RICARDO VALENCIA RAMÍREZ
Subdirector

MARÍA FERNANDA DE LA OSSA ARCHILA
Secretaria General

DIRECTORES TÉCNICOS

ANTONIO JOSÉ AVENDAÑO AROSEMENA
Dirección de Metodología y Producción
Estadística

LILIANA ACEVEDO ARENAS
Dirección de Censos y Demografía

JULIETH ALEJANDRA SOLANO VILLA
Dirección de Regulación, Planeación,
Estandarización y Normalización

JOVANA PALACIOS MATALLANA
Dirección de Síntesis y Cuentas Nacionales

SANDRA LILIANA MORENO MAYORGA
Dirección de Geoestadística

MAURICIO ORTIZ GONZÁLEZ
Dirección de Difusión, Mercadeo y Cultura
Estadística

© DANE, 2020

Prohibida la reproducción total o parcial sin
permiso o autorización del Departamento
Administrativo Nacional de Estadística,
Colombia.

LUISA FERNANDA SUÁREZ LEÓN
Coordinadora de Regulación

RAFAEL HUMBERTO ZORRO CUBIDES
JOSÉ ANDERSON CASTAÑEDA ZAMORA
Coordinación de Regulación

SONIA MARCELA NARANJO MORALES
Corrección de estilo

CONTENIDO

Introducción	4
1. El software Nesstar Publisher y los estándares estadísticos: Data Documentation Initiative (DDI) y Dublin Core (DC)	6
1.1 Nesstar Publisher	7
1.2 Data Documentation Initiative (DDI)	7
1.3 Dublin Core (DC)	7
1.4 Ventajas del uso de los estándares estadísticos DDI y DC	8
2. Proceso de documentación	9
2.1 Descripción del documento	14
2.2 Descripción de la operación estadística	17
2.3 Base de datos	46
2.4 Grupo de variables	58
2.5 Materiales de referencia externos	67
3. Archivo Nacional de Datos (ANDA)	76
Bibliografía	77

Introducción

Para incrementar el valor de la información estadística y potenciar su uso en la toma de decisiones en política pública, por parte de la academia, el sector privado y la sociedad, es relevante fortalecer los procesos de documentación metodológica en el Sistema Estadístico Nacional (SEN). Para tal fin se han desarrollado iniciativas internacionales impulsadas por el Banco Mundial y la Organización para la Cooperación y el Desarrollo Económico (OECD), como: el Archivo Nacional de Datos (ANDA) y los estándares empleados para la documentación y la difusión de metadatos y microdatos, como el Data Documentation Initiative (DDI), el Dublin Core (DC) y el Statistical Data and Metadata Exchange (SDMX), que se enmarcan en el Programa Acelerado de Datos (PAD).

Estas iniciativas facilitan la aplicación de las mejores prácticas y los estándares internacionales sobre documentación y difusión de metadatos y microdatos para fortalecer y estandarizar la documentación, la preservación y la difusión de los metadatos asociados a una operación estadística. Además, estas son complementadas con el Plan Estadístico Nacional 2017 – 2022 y la Norma Técnica de la Calidad del Proceso Estadístico (NTCPE 1000:2017), que establecen y promueven acciones referentes a la implementación de lineamientos y estándares definidos por el DANE y demás referentes metodológicos internacionales, de tal forma que se generen resultados coherentes y comparables a nivel nacional e internacional.

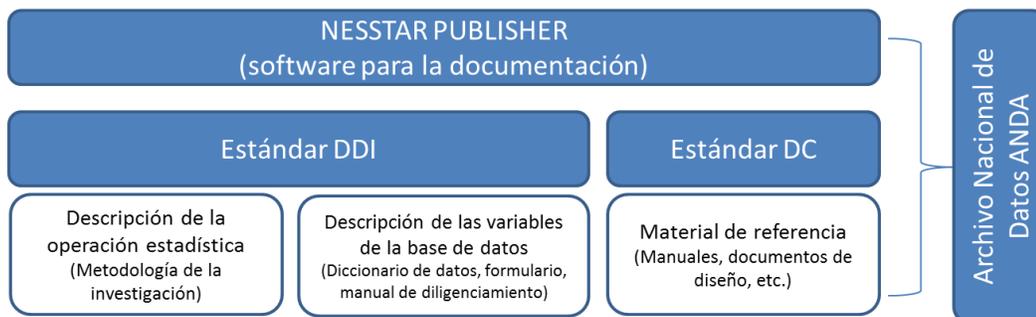
Bajo este escenario, el DANE como ente coordinador y regulador del SEN encargado de dar lineamientos en materia de producción y difusión, elaboró la Guía para la implementación de los estándares Data Documentation Initiative (DDI) y Dublin Core (DCMI), con el fin de presentar los pasos a seguir para documentar encuestas, censos y operaciones estadísticas basadas en registros administrativos.

1. El software Nesstar Publisher y los estándares estadísticos: Data Documentation Initiative (DDI) y Dublin Core (DC)

El DANE consciente de la necesidad de brindar a los usuarios mayor claridad, y confianza técnica y en cumplimiento de los principios de Transparencia y Accesibilidad en el proceso de las operaciones estadísticas que se desarrollan en el SEN, se vincula en el 2008 a la iniciativa internacional Programa Acelerado de Datos (PAD), cuyo propósito ha sido documentar, difundir y preservar microdatos de operaciones estadísticas, conforme a estándares y prácticas internacionales, así como brindar parámetros para la presentación de los documentos técnicos y metodológicos utilizados por cada operación estadística.

El PAD fue lanzado en 2006 como recomendación del Plan de Acción para Estadísticas de Marrakech MAPS. Fue concebido como instrumento de fortalecimiento de las estadísticas empleadas para dar seguimiento a los Objetivos de Desarrollo del Milenio, hoy Objetivos de Desarrollo Sostenibles (ODS), así como para fortalecer la producción estadística y el desarrollo y la implementación de programas de divulgación y promoción. Además, se ha centrado en apoyar la implementación de las mejores prácticas y las normas internacionales aprovechando la labor realizada por la Red Internacional de Encuestas a Hogares (IHSN - International Household Survey Network) y armonizar tanto las prácticas de recopilación como el análisis de datos en aras de promover la comparabilidad de la información estadística.

Figura 1. Usos de los estándares estadísticos DDI y DC



Fuente: DANE

Para su implementación en operaciones estadísticas por censo, por muestreo y a partir de registros administrativos, se adaptaron como estándares de documentación el Data Documentation Initiative (DDI) y el Dublin Core (DC), cuya aplicación requiere de la utilización del software especializado de uso libre Nesstar Publisher y cuyos resultados son visibles a través de un portal centralizado denominado Archivo Nacional de Datos (ANDA).

1.1 Nesstar Publisher

Es el software editor para la documentación de los metadatos y los microdatos para su publicación en el ANDA¹; funciona como plantilla para la documentación que cumple con los estándares DDI y DC. Es un software libre y gratuito para la creación de metadatos.

Los metadatos proporcionan un puente entre los productores de datos y sus usuarios, pues transmiten información que es esencial para el análisis secundario y facilitan la interpretación de los resultados.

1.2 Data Documentation Initiative (DDI)

La iniciativa de documentación de datos se origina en 1995 como un esfuerzo internacional para establecer un estándar basado en lenguaje XML, cuyo propósito es la documentación de los metadatos y los microdatos. Proporciona un instrumento simple para registrar y comunicar todas las características sobresalientes de los mismos.

La plantilla IHSN_1.6_es_aj_20160301 de la Red Internacional de Encuestas de Hogares, estandarizada de acuerdo con el estándar DDI, se utiliza para documentar encuestas, censos y operaciones estadísticas basadas en registros administrativos.

Una vez configurada la plantilla en el software Nesstar Publisher, el usuario puede realizar: la documentación de la descripción del documento, la descripción de la operación estadística y el cargue de la base de datos.

1.3 Dublin Core (DC)

La plantilla IHSN_1.5_ del estándar Dublin Core (DC) se utiliza para documentar material relacionado como manuales, guías, instructivos y demás documentos o reportes generados durante el desarrollo de la operación estadística.

Una vez configurada la plantilla del Nesstar Publisher, el usuario puede realizar el enlace y la documentación de metodologías, fichas metodológicas, cuestionarios y otros documentos que el usuario considere de importancia para la operación estadística.

Para incluir la documentación en cada uno de los campos, la herramienta proporciona una descripción del elemento, como ejemplo de lo que se puede incluir en el campo correspondiente.

¹ Para más información: http://formularios.dane.gov.co/Anda_4_1/index.php/home

1.4 Ventajas del uso de los estándares estadísticos DDI y DC

- Permiten el uso de un vocabulario estándar para documentar los datos de la investigación y conducen a la coherencia y a la interoperabilidad.
- Están diseñados para que los datos de la operación estadística sean comprensibles de forma independiente.
- Proporcionan una estructura estándar para todos los metadatos que acompañan a un conjunto de datos y ayuda a los usuarios y analistas de datos a interpretar su contenido.
- Promueven la interoperabilidad entre los sistemas de información y de los procedimientos a los que estos dan soporte, lo que facilita el intercambio de datos y metadatos.
- Es un estándar abierto y no propietario por lo que cualquiera puede usarlo. Puede servir como una base para los catálogos de datos, ya que proporciona una estructura estándar para la búsqueda en el estudio y los niveles variables para los usuarios.
- Permiten la reutilización de metadatos: "introduzca una vez, use a menudo" significa que puede reutilizar metadatos durante el curso del ciclo de vida de los datos para evitar la duplicidad de esfuerzos.

2. Proceso de documentación

Para iniciar el proceso de documentación es importante contar con los siguientes elementos:

- Identificación de actores.
- Información para el proceso de documentación:
 - La metodología de la operación estadística.
 - El cuestionario de preguntas o el registro en el que se basa la operación estadística.
 - El material de referencia empleado para complementar la información consignada en el metadato: manuales, guías o instructivos.

A continuación, se describen cada uno de los elementos que intervienen en el proceso de documentación.

a) Identificación de actores

En el proceso de documentación de metadatos interviene el coordinador y temático responsables de la operación estadística y que tienen el conocimiento y la responsabilidad sobre el proceso estadístico. Un actualizador responsable de verificar la información documentada en el Nesstar Publisher y el responsable de realizar la publicación en el Archivo Nacional de Datos (ANDA).

b) Información para el proceso de documentación

Metodología

Debe comprender los procedimientos y las técnicas estadísticas para el diseño de la obtención de los datos y el análisis de estos, así como, la interpretación y la representación de los resultados en forma apropiada.

Bases de datos

Se debe contar con la siguiente información:

- **Bases de datos:** conjunto o colección de datos interrelacionados entre sí, que se utilizan para la obtención de información de acuerdo con el contexto de estos y que son almacenados sistemáticamente para su posterior uso, Contiene las variables y los registros de una operación estadística. En aras de asegurar la confidencialidad de las fuentes debe surtir un proceso de anonimización de las bases de datos² previo a su difusión a través del ANDA.
- **Diccionario de datos:** contiene información de la base de datos que puede ser el nombre de la tabla o el archivo de donde viene la variable, el nombre del campo, la descripción del campo, el tipo y la medida del dato y la longitud del campo.

Material de referencia

Corresponde a aquellos documentos que hacen parte del diseño de la operación estadística o aquellos que complementan la información contenida en el metadato. Entre estos se incluyen:

- **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.

Adicionalmente:

- Ficha metodológica
- Cuestionario
- Presentaciones.

² Para más información:

http://www.sen.gov.co/files/RegistrosAdministrativos/Gu%C3%ADa_para_la_anonimizaci%C3%B3n_de_bases_de_datos.pdf

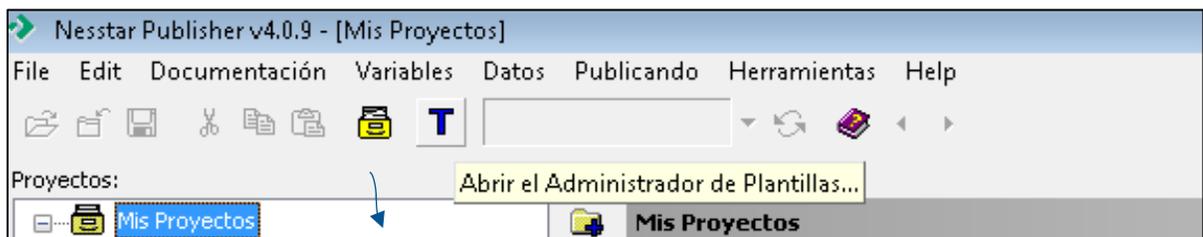
c) Herramientas de Gestión de Microdatos

Para iniciar el proceso de documentación, se requiere descargar³ las siguientes herramientas de gestión de microdatos:

- Plantilla DDI para censos y encuestas por muestreo: IHSN_1.6_es_aj.NesstarTemplate.
- Plantilla DDI para uso estadístico de registros administrativos: IHSN_1.6_es_aj_RRAA.NesstarTemplate.
- Plantilla Dublin Core para censos, encuestas por muestreo y uso estadístico de registros administrativos: IHSN_1.5_es.NesstarRDTemplate.
- El software Nesstar Publisher a través del que se desarrolla la documentación.

Este es el proceso de documentación en el software Nesstar Publisher:

- Se clasifica la información a documentar.
- Instalar el Nesstar Publisher e incorporar las plantillas dando clic sobre el administrador de plantillas identificado con la letra T, se despliegan dos opciones de plantillas:
 1. Plantillas de operación estadística.
 2. Plantillas de descripción de los materiales de referencia.

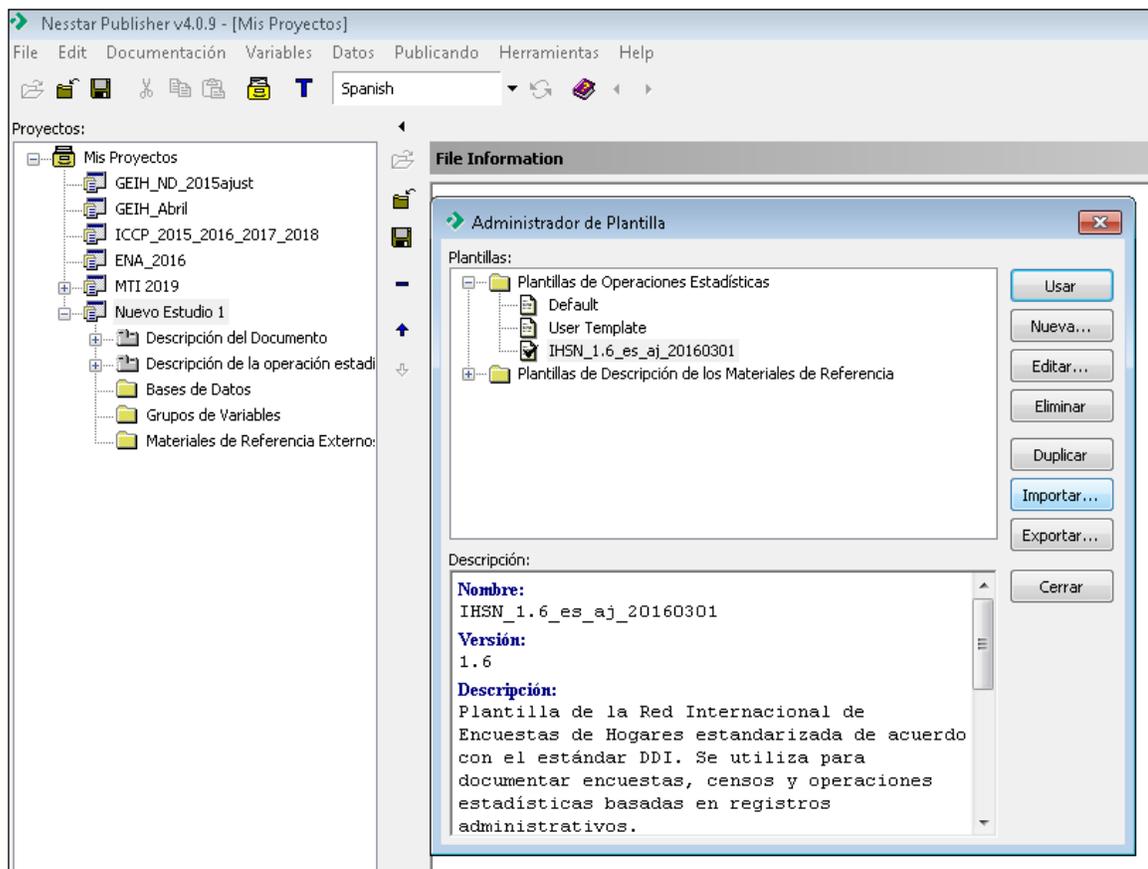


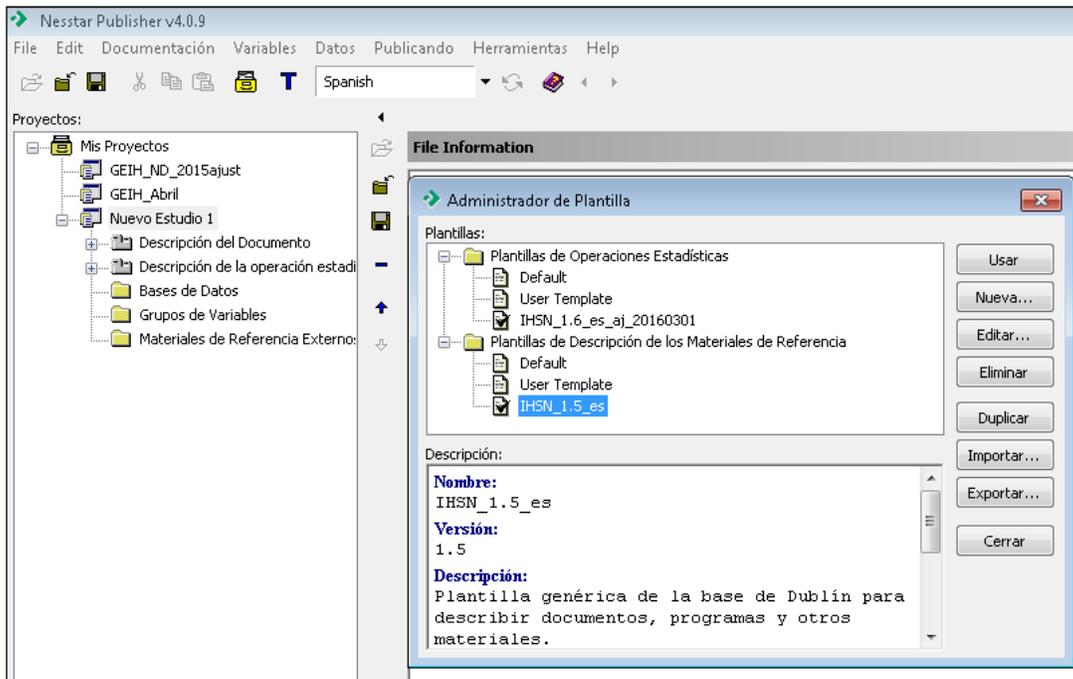
En la opción plantillas de operaciones estadísticas se puede seleccionar: la plantilla **IHSN_1.6_es_aj.NesstarTemplate** que se usa para documentar operaciones estadísticas por censos

³ Estas herramientas se encuentran disponibles en la página del ANDA: <http://andacna.dane.gov.co/index.php/herramientas>

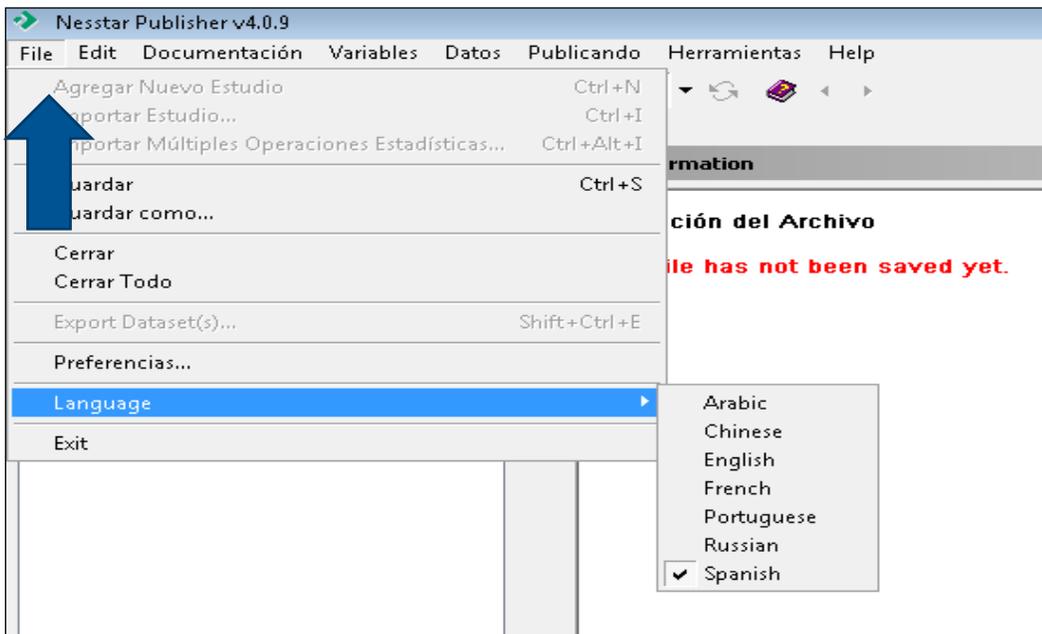
o encuestas por muestreo, o la *plantilla DDI* para uso estadístico de registros administrativos **IHSN_1.6_es_aj_RRAA.NesstarTemplate**.

- Se activa la plantilla, después de importada seleccionando la opción **Usar**, debe aparecer un visto bueno sombreado.
- En la opción plantillas de descripción de los materiales de referencia se selecciona e importa la plantilla IHSN_1.5_es.
- Para finalizar se selecciona **Cerrar**.

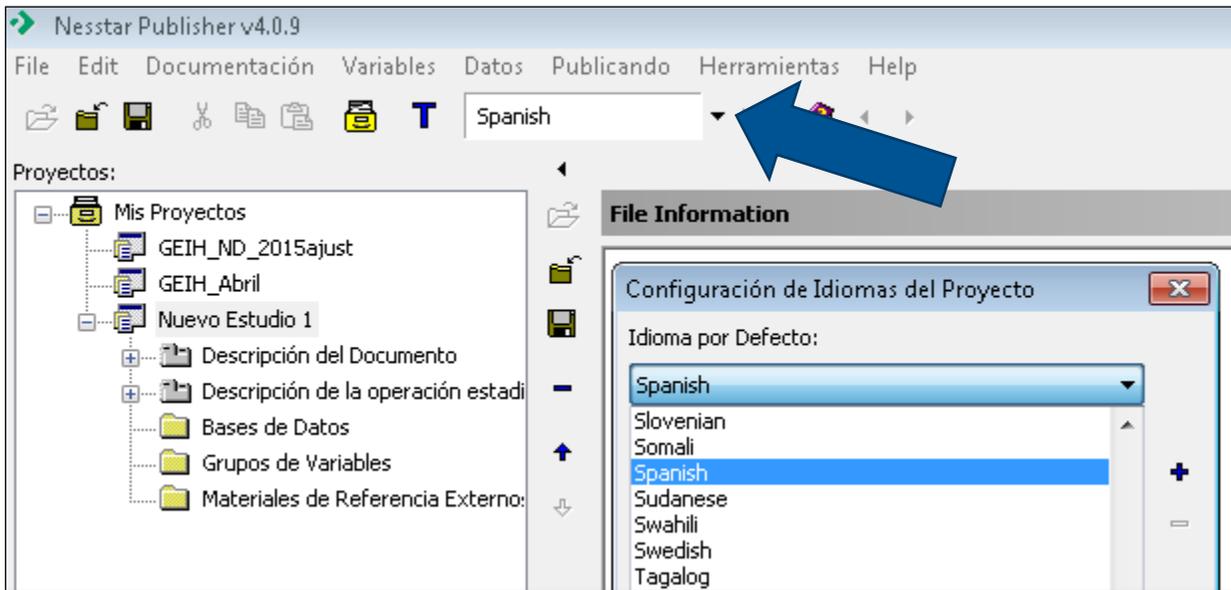




- Se selecciona el idioma con las opciones **File** y luego **Language**, spanish.



- Se configura el idioma del proyecto.

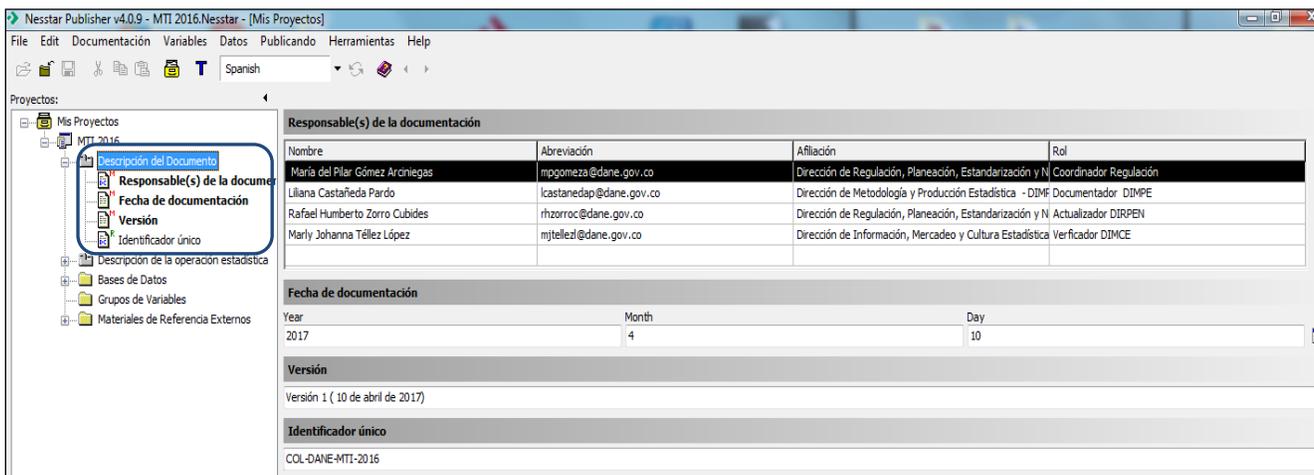


Luego se realiza la documentación de las siguientes secciones: descripción del documento, descripción de la operación estadística, base de datos y material de referencia.

2.1 Descripción del documento

Es importante proveer en el metadato la información sobre las personas y las entidades que intervinieron en la operación estadística y en el proceso de documentación.

Gráfica 1. Descripción del documento

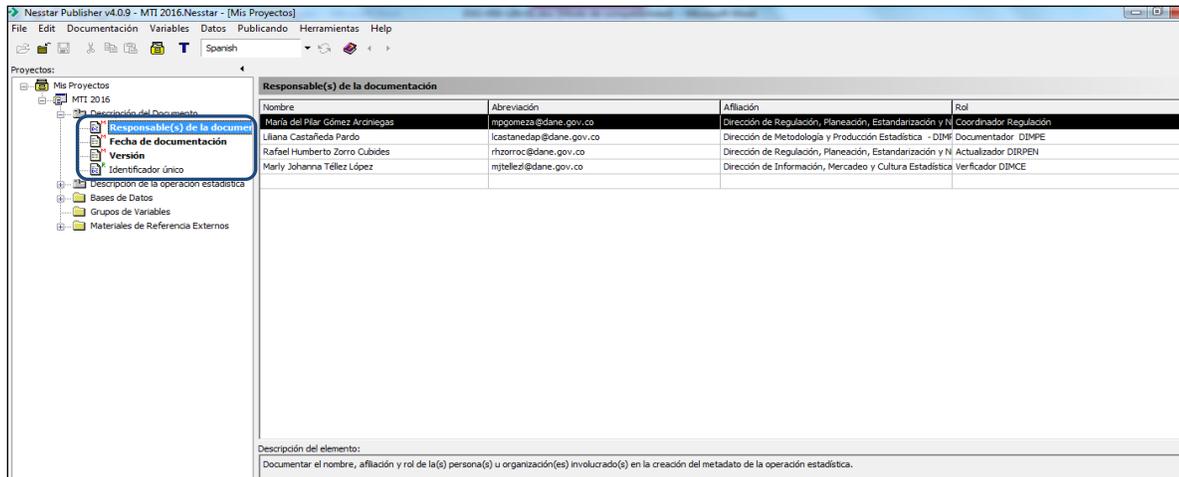


Fuente: DANE.

La sección está compuesta por los siguientes campos:

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

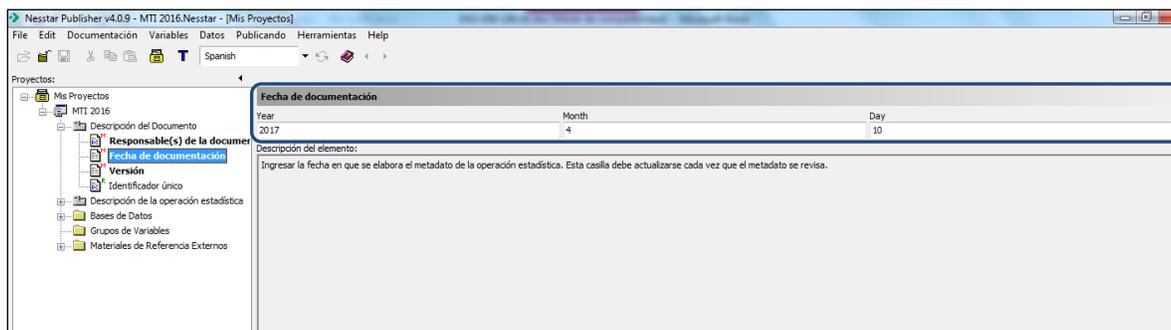
Gráfica 2. Responsables de la documentación



Fuente: DANE.

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

Gráfica 3. Fecha de documentación



Fuente: DANE.

Versión: esta opción permite identificar la versión del documento y describir los cambios relevantes que contiene la nueva versión.

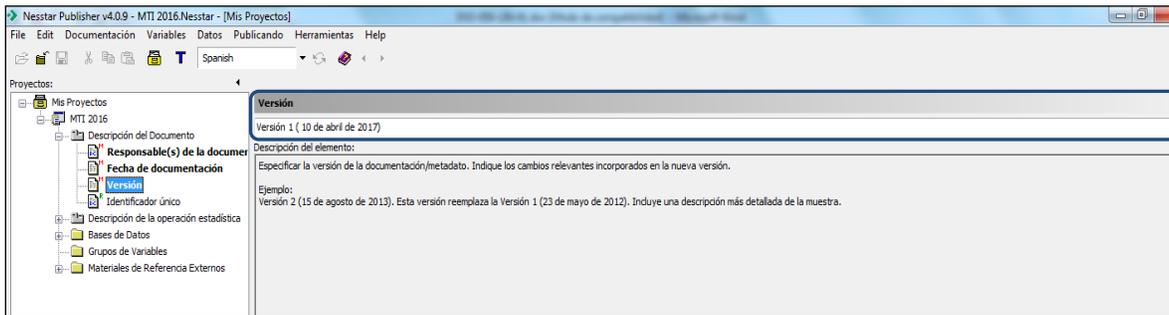
Ejemplo:

Versión 1 (enero de 2017).

Versión 2 (febrero de 2017).

Esta versión reemplaza la Versión 1 (enero de 2017) ya que incluye una descripción más detallada de la muestra.

Gráfica 4. Campo Versión



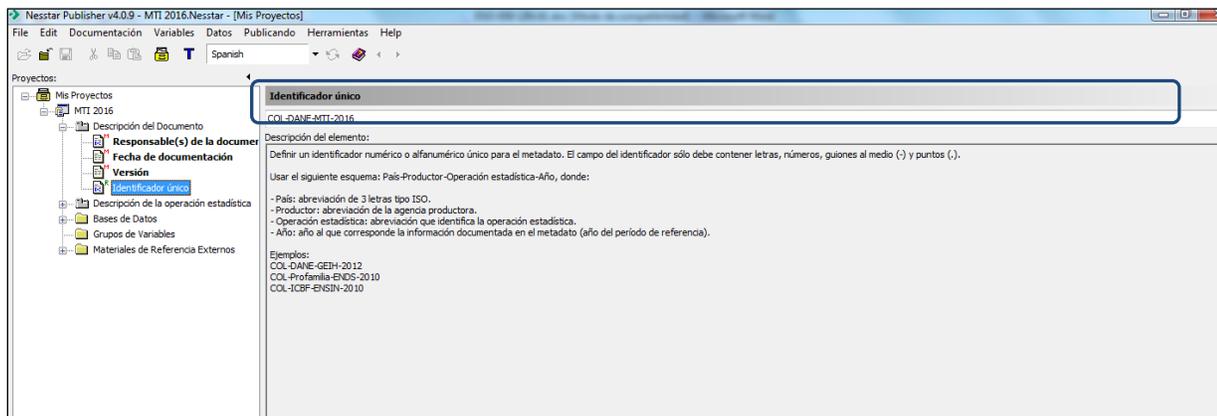
Fuente: DANE.

Identificador único: corresponde al código único de identificación del metadato, el cual puede ser numérico o alfanumérico.

Usa el esquema país-productor-operación estadística-año, donde:

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

Gráfica 5. Identificador único

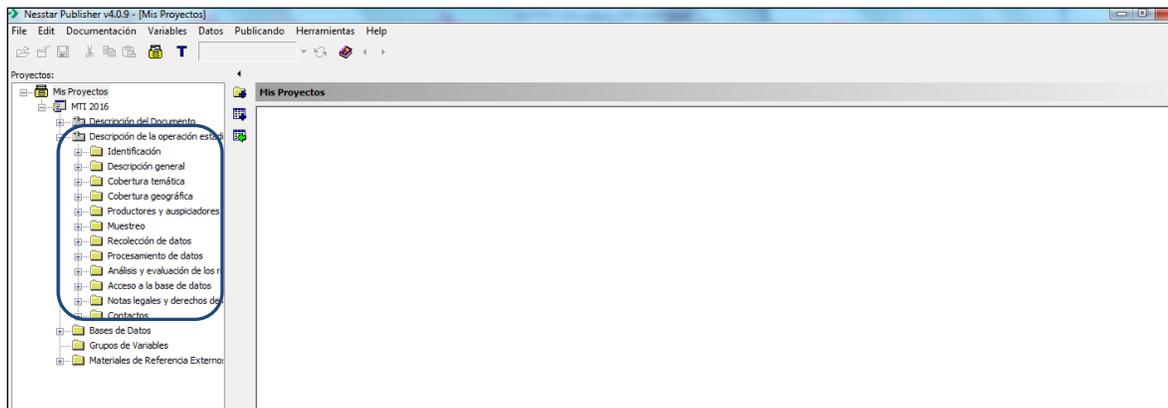


Fuente: DANE.

2.2 Descripción de la operación estadística

Esta sección contiene doce campos relacionados con la información general de la operación estadística como: identificación, descripción general, cobertura temática, cobertura geográfica, productores y auspiciadores, muestreo recolección de datos, procesamiento de datos, análisis evaluación, Acceso a la base de datos, notas legales y derechos de autor y contactos.

Gráfica 6. Descripción de la operación estadística

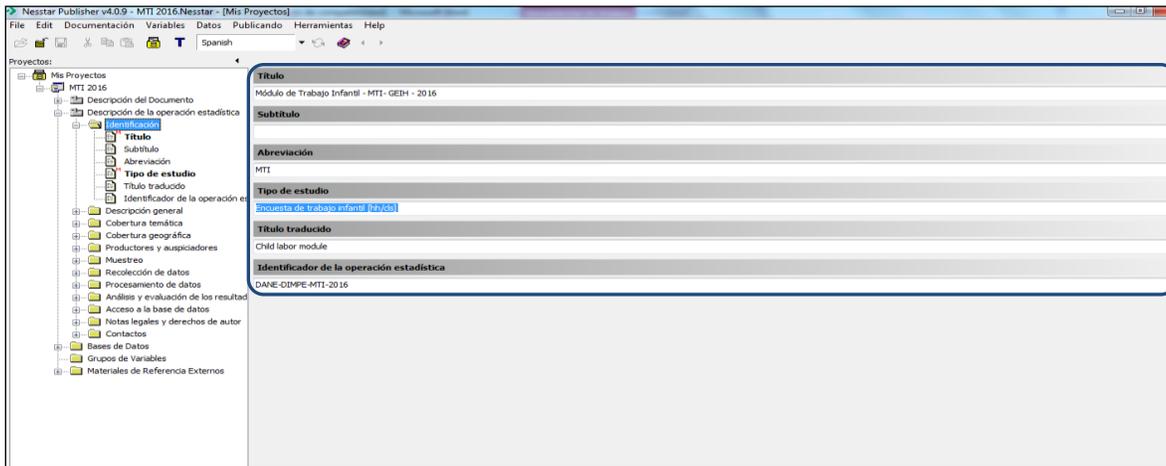


Fuente: DANE.

IDENTIFICACIÓN

Corresponde a un grupo de campos relacionados con información de identificación de la operación estadística como: título, subtítulo, abreviación, tipo de estudio, título traducido e identificador de operación estadística.

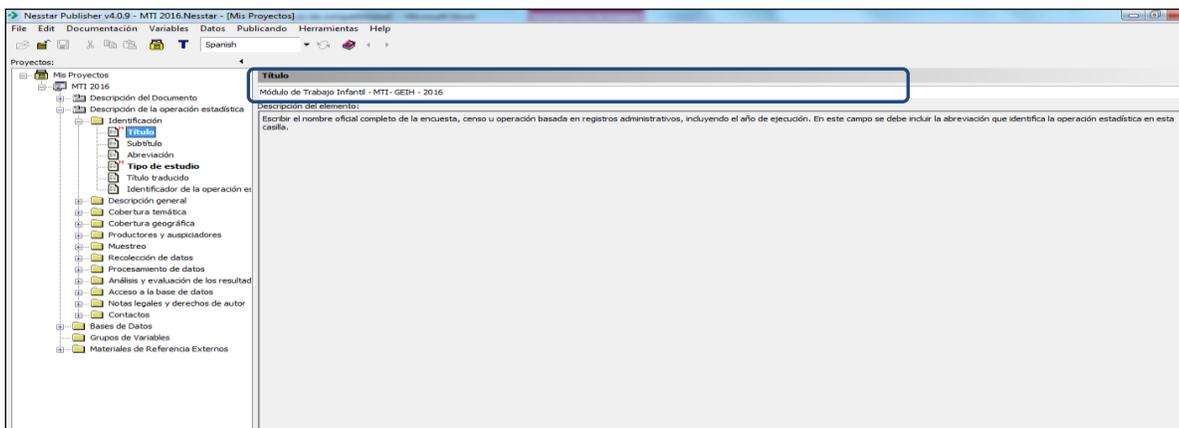
Gráfica 7. Identificación



Fuente: DANE.

Título: hace referencia al nombre completo de la operación estadística, incluyendo la sigla de la operación estadística y el año de ejecución.

Gráfica 8. Título



Fuente: DANE.

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

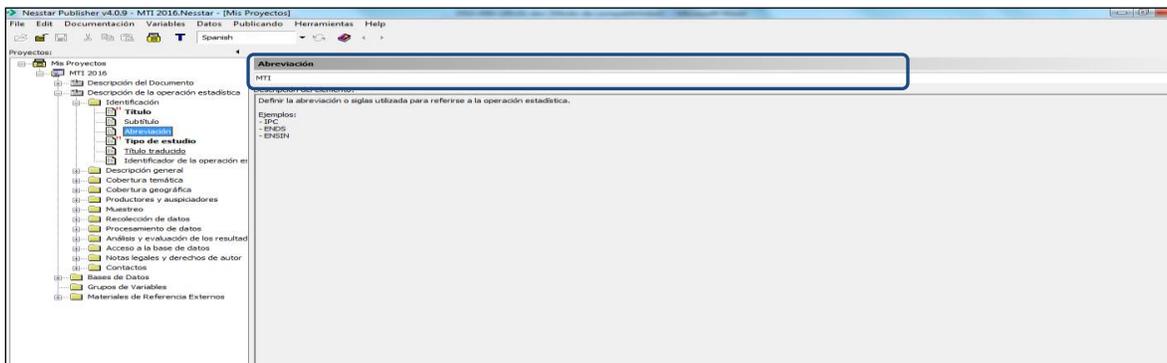
Gráfica 9. Subtítulo



Fuente: DANE.

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

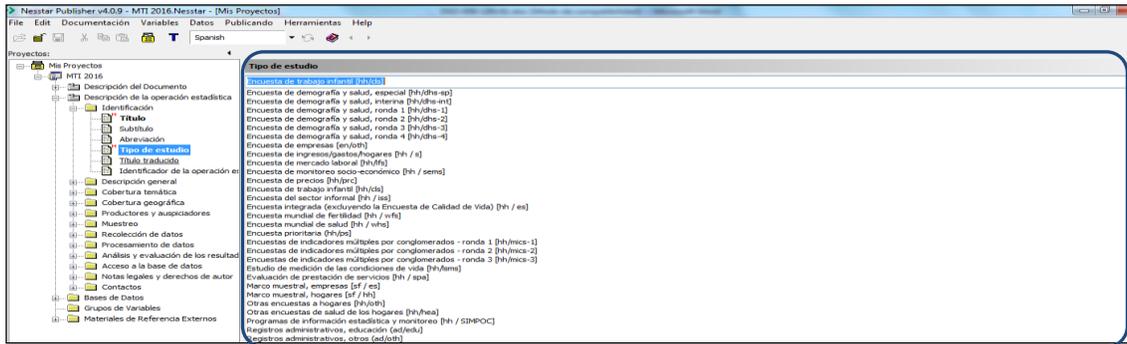
Gráfica 10. Abreviación



Fuente: DANE.

Tipo de estudio: se selecciona de acuerdo con la temática y al método de recolección o acopio de la lista desplegable. Si ninguna de las opciones listadas es apropiada, se puede incorporar la categoría correspondiente con una nueva entrada en el editor.

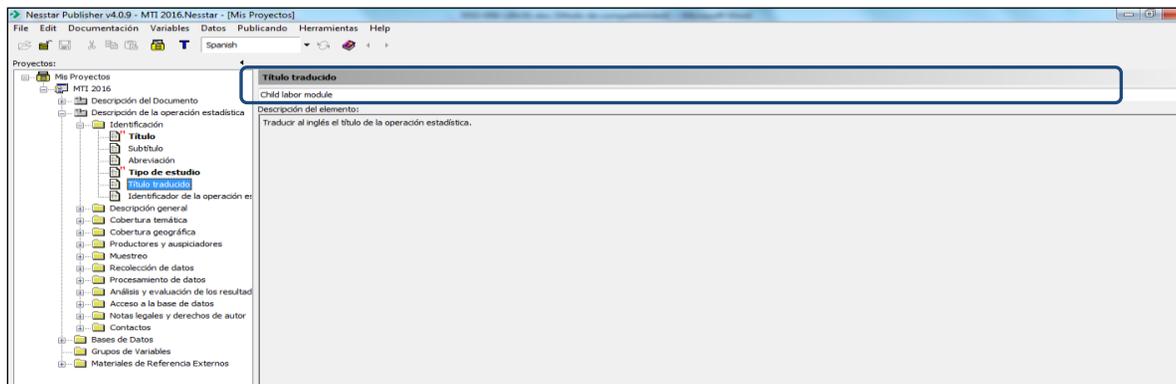
Gráfica 11. Tipo de estudio



Fuente: DANE.

Título traducido: se coloca el título de la operación estadística traducido al inglés. Se sugiere colocar la primera letra en mayúscula, como, por ejemplo: *Sample monthly hotels*.

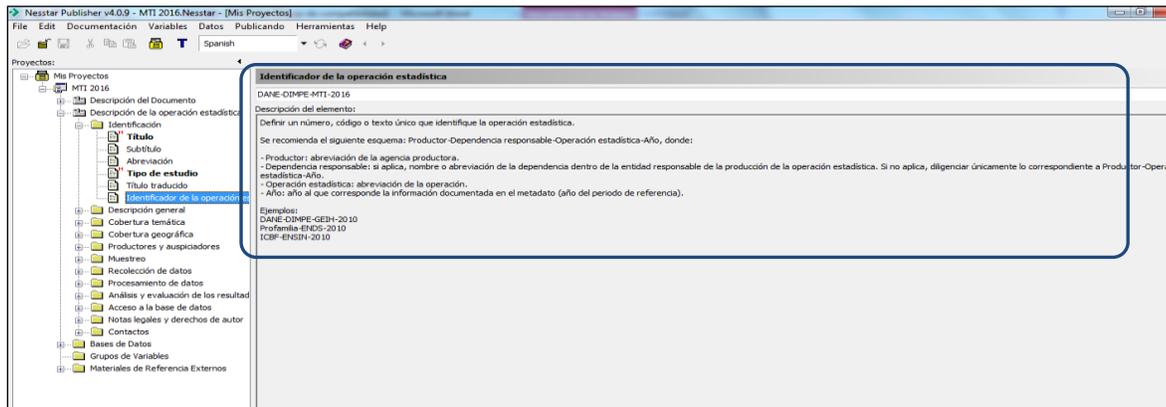
Gráfica 12. Título traducido



Fuente: DANE.

Identificador de la operación estadística: corresponde al número o el 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 (2016).

Gráfica 13. Identificador de la operación estadística

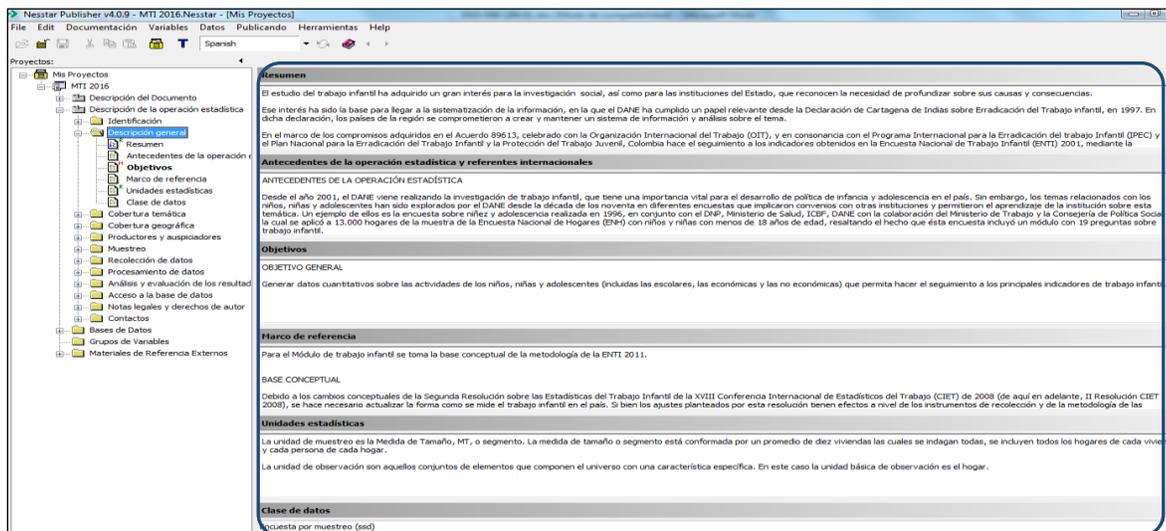


Fuente: DANE.

DESCRIPCIÓN GENERAL

Corresponde a un grupo de campos donde se relaciona información de la operación estadística: resumen, antecedentes de la operación estadística, objetivos, marco de referencia, unidades estadísticas y clase de datos.

Gráfica 14. Descripción general



Fuente: DANE.

Resumen: esta sección presenta de forma clara, ordenada y resumida una visión general del alcance, la importancia y las implicaciones de la operación estadística.

Antecedentes de la operación estadística y referentes internacionales: se describe el origen, la evolución histórica de la operación estadística, se señalan sus principales hitos y las 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, lo cual permite construir una visión comparativa acerca de los cambios metodológicos que pueden incidir en la obtención y el análisis de sus resultados.

Asimismo, se relacionan los referentes internacionales de la operación estadística como es el caso de los organismos internacionales como ONU, EUROSTAT, institutos nacionales de estadística reconocidos a nivel mundial y presenta las principales recomendaciones adoptadas y/o adaptadas. En el caso de nuevas operaciones, resulta importante contextualizar al usuario sobre el origen de esta, referentes de estudios que aborden temáticas o que cuenten con características similares y presentar las conclusiones más relevantes de las pruebas piloto realizadas en el diseño de la operación estadística.

Objetivos: se relaciona el objetivo general y los objetivos específicos que se establecieron para la operación estadística y la finalidad del tema de investigación.

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.

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

Marco conceptual: describe, explica y establece las relaciones entre los conceptos fundamentales de la operación estadística. Corresponde a los conceptos o los términos técnicos básicos utilizados los cuales deben ser estandarizados o armonizados juntamente con el DANE, teniendo en cuenta referentes internacionales. Los demás términos que se empleen se relacionan en el glosario, en el campo «Palabra(s) clave».

Marco legal: describe la normatividad en la que se circunscribe la operación estadística.

Unidades estadísticas: se describe la unidad de análisis básico (individuos, hogares, establecimientos, empresas, etc.). Es la unidad que proporciona la base para los agregados estadísticos y a los que se refiere los datos tabulados, comprende la unidad de observación⁴, la unidad de análisis⁵ y la unidad de muestreo⁶. Cada una de las unidades del estudio a documentar se identifica con el título en mayúscula.

⁴ Unidad de observación: elemento o conjunto de elementos sobre los que se hace la medición de las diferentes variables en una operación estadística.

⁵ Unidad de análisis: elemento de estudio sobre el que se presentan los resultados o las conclusiones de la operación estadística.

⁶ Unidad de muestreo: elemento o conjunto de elementos definidos o establecidos a ser seleccionados en la muestra.

Ejemplo:

UNIDAD DE OBSERVACIÓN

Fincas productoras de flores bajo invernadero y a cielo abierto.

UNIDAD DE ANÁLISIS

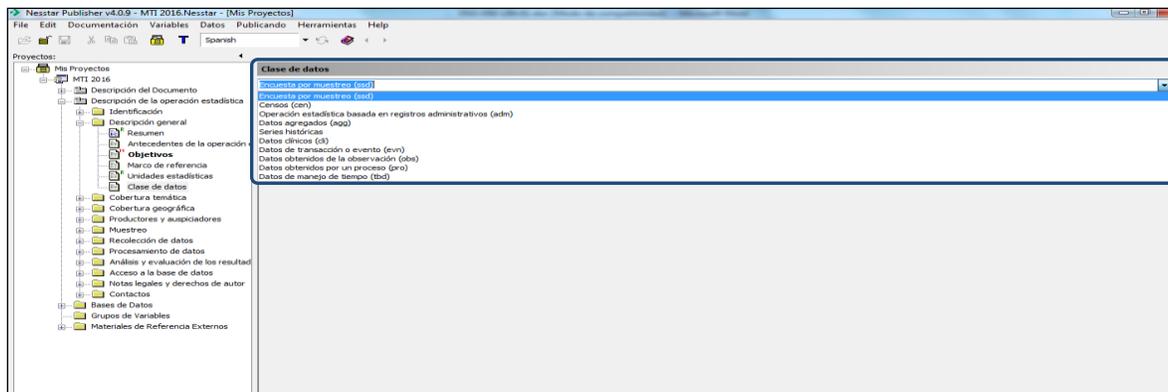
Fincas y lotes dedicados al cultivo de flores bajo invernadero y a cielo abierto.

UNIDAD DE MUESTREO

Fincas productoras de flores bajo invernadero y a cielo abierto.

Clases de datos⁷: se selecciona de la lista desplegable el tipo de datos u operación estadística (encuesta por muestreo, censo, operación estadística a partir de registros administrativos), recolectados durante el estudio.

Gráfica 15. Clase de datos



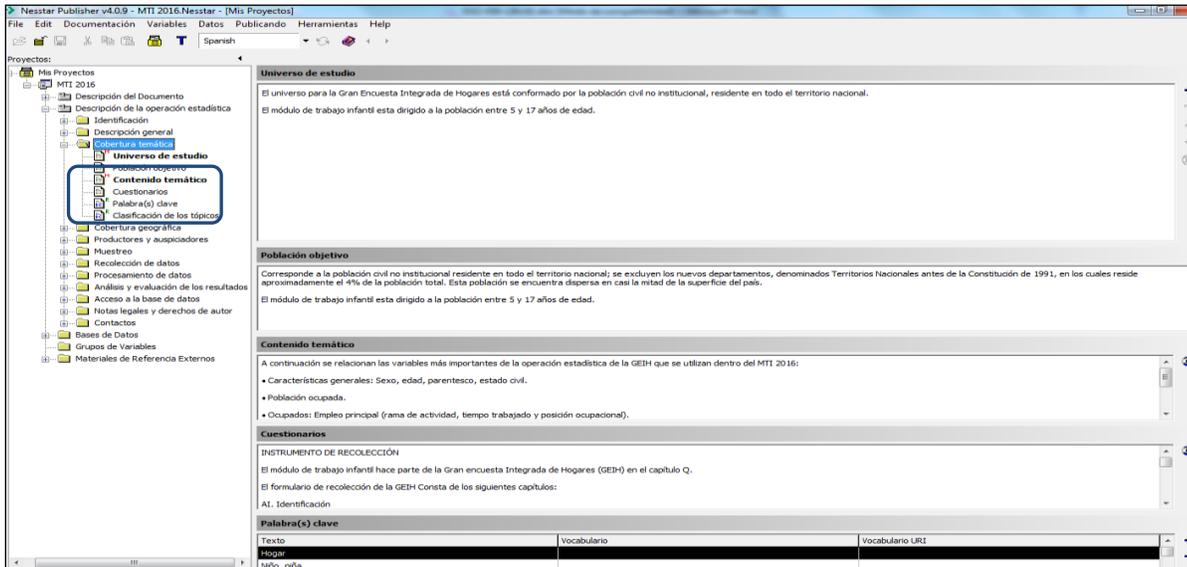
Fuente: DANE.

COBERTURA TEMÁTICA

Está conformado por seis campos donde se relaciona información relacionada con el universo de estudio, población objetivo, contenido temático, cuestionarios, palabra(s) clave(s) y clasificación de tópicos.

⁷ Hace referencia a la metodología estadística.

Gráfica 16. Cobertura temática



Fuente: DANE.

Universo de estudio: describe el conjunto de unidades o individuos a los que se refiere el estudio y que satisfacen una definición común. Refleja los objetivos del estudio, por cuanto los describe en términos de contenido, unidades, espacio y tiempo.

Población objetivo: describe el conjunto de elementos definidos en un espacio y en un tiempo sobre los que se recolectó información. Esta descripción incluye un listado de todos los grupos excluidos de la operación estadística, y de ser relevante, el motivo por el cual se excluyen. Al igual que los anteriores elementos los títulos deben ir en mayúscula.

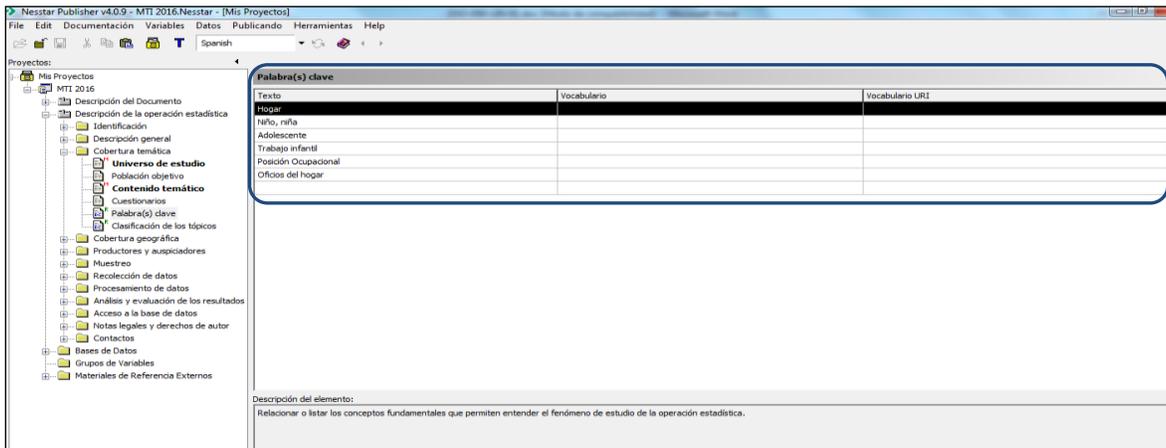
Contenido temático: menciona las variables (o bloques de variables) más importantes de la operación estadística. Cuando se cuenta con un gran número de variables se hace necesario clasificarlas por los principales temas y para diferenciar cada tema, los títulos deben quedar en mayúscula.

Cuestionarios: describe la estructura y la composición del cuestionario, se documenta los objetivos de cada uno de los módulos o los capítulos, así como los objetivos de las preguntas.

Nota: por lo general, las operaciones estadísticas basadas en registros administrativos no cuentan con cuestionario.

Palabra (s) clave: relaciona o lista los conceptos fundamentales que permiten entender el fenómeno de estudio de la operación estadística. Su objetivo es facilitar la búsqueda y la clasificación de los proyectos archivados.

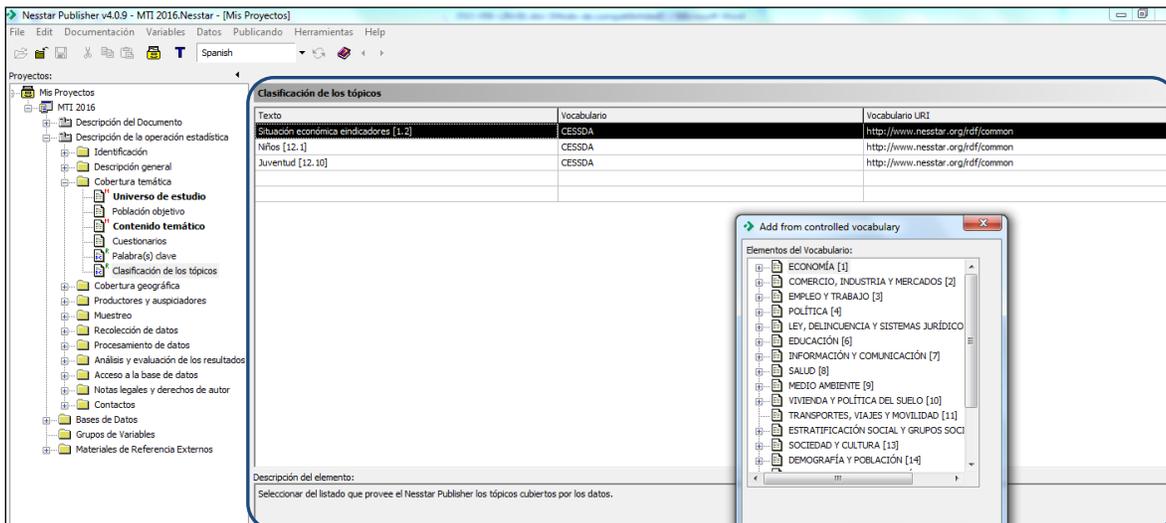
Gráfica 17. Palabra(s) clave



Fuente: DANE.

Clasificación de los tópicos: se selecciona de la lista desplegable los tópicos cubiertos por los datos. El atributo «vocabulario URI» especifica la ubicación del vocabulario controlado.

Gráfica 18. Clasificación de los tópicos

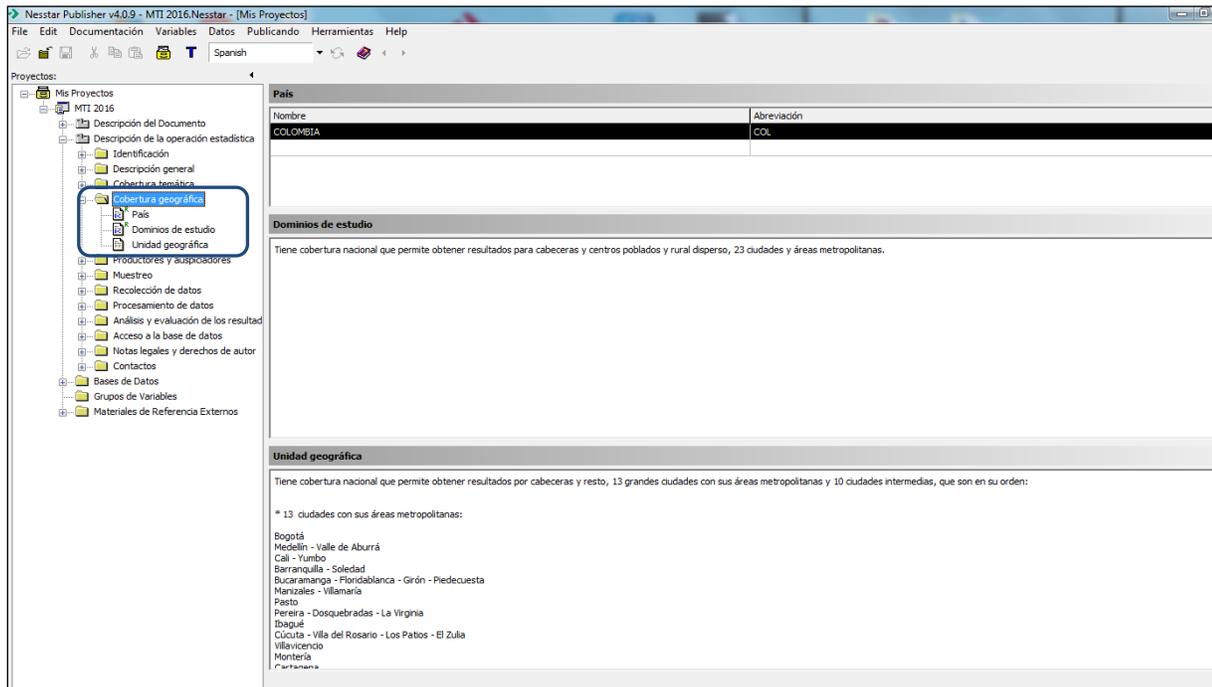


Fuente: DANE.

COBERTURA GEOGRÁFICA

Está conformado por tres campos a saber: país, dominios de estudio y unidad geográfica.

Gráfica 19. Cobertura geográfica



Fuente: DANE.

País: se selecciona de la lista desplegable el país donde se ejecuta la operación estadística. Se pone el nombre del país en mayúscula y su respectiva abreviación en código ISO que corresponde a las tres primeras letras.

Gráfica 20. País



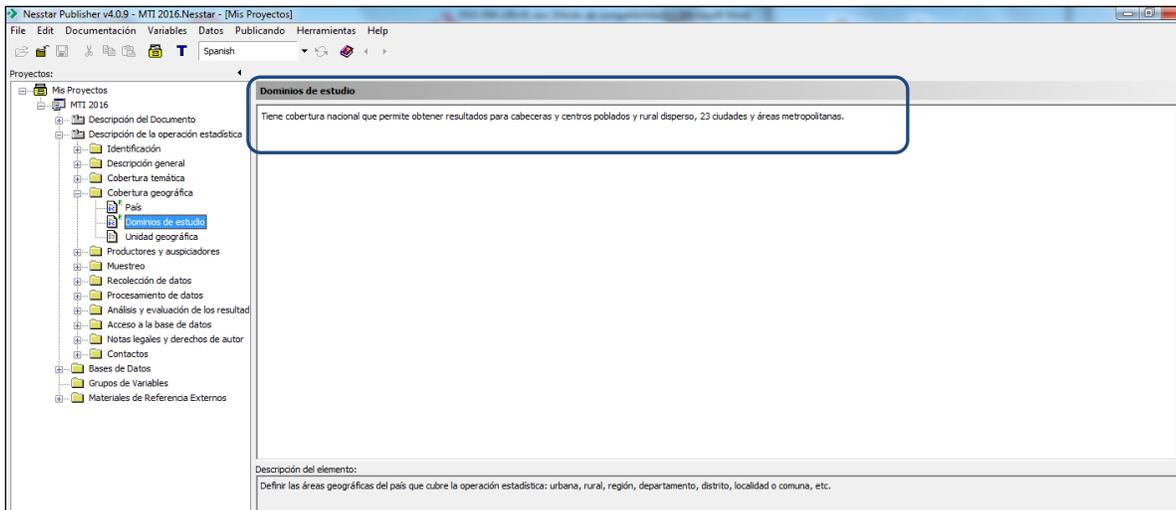
Fuente: DANE.

Dominios de estudio: describe el subgrupo de la población objetivo sobre el que se presentan resultados, puede ser 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 denominados *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, Glossary).

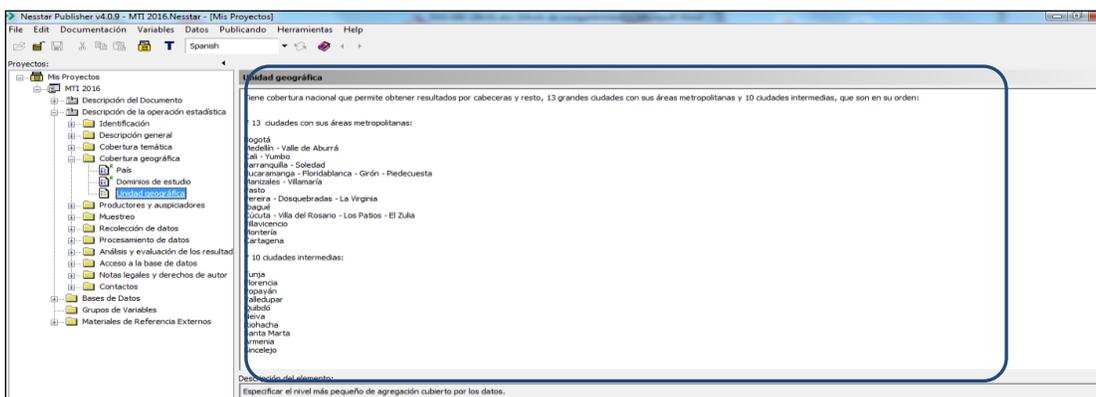
Gráfica 21. Dominios de estudio



Fuente: DANE.

Unidad geográfica: se proporciona información acerca de las áreas que fueron cubiertas por la operación estadística, se pueden incluir las razones por las que la encuesta no cubrió el área total del país. Se especifica el nivel más pequeño de agregación cubierto por los datos.

Gráfica 22. Unidad geográfica



Fuente: DANE.

PRODUCTORES Y AUSPICIADORES

En este campo se relacionan los nombres de las personas y entidades responsables del diseño, la implementación y el financiamiento de la operación estadística. Se desagrega en cuatro campos investigador principal, colaboradores, financiamiento y reconocimientos.

Gráfica 23. Productores y auspiciadores

Investigador principal			
Nombre	Afilación		
Director de Metodología y Producción Estadística - DIMPE			
Colaboradores			
Nombre	Abreviación	Afilación	Rol
Director de Metodología y Producción Estadística - DIMPE - DIMPE - GDSH		Departamento Administrativo Nacional de Estadística - DA Equipo Técnico	
Financiamiento			
Agencia/Organización	Abreviación	Número de Donación	Rol
Departamento Administrativo Nacional de Estadística	DANE		Ejecutor
Reconocimientos			
Nombre	Afilación		Rol

Fuente: DANE.

Investigador principal: relaciona el responsable del diseño y la dirección de la operación estadística. Cabe recordar que las operaciones estadísticas pueden tener más de una persona o entidad responsable. También se puede poner el nombre del área productora.

Gráfica 24. Investigador principal

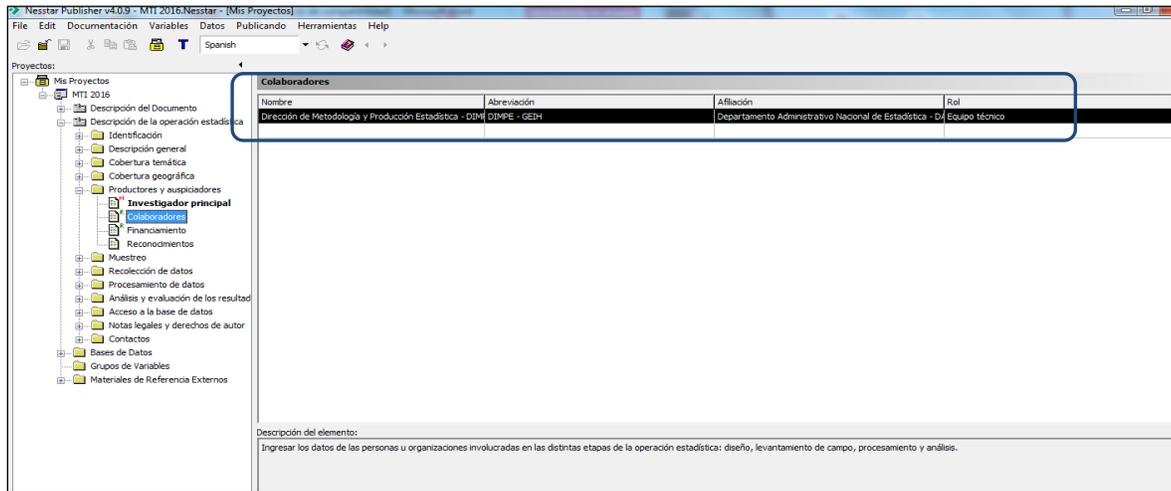
Investigador principal	
Nombre	Afilación
Director de Metodología y Producción Estadística - DIMPE	

Descripción del elemento:
Ingresar los datos de la unidad o persona responsable del diseño y conducción de la operación estadística.

Fuente: DANE.

Colaboradores: relaciona las personas y las organizaciones involucradas en las distintas etapas de la operación estadística (diseño, levantamiento de campo, procesamiento y análisis), identificando el rol.

Gráfica 25. Colaboradores



The screenshot shows the 'Colaboradores' table with the following data:

Nombre	Abreviación	Afiliación	Rol
Dirección de Metodología y Producción Estadística - DIM	DIMPE - GEH	Departamento Administrativo Nacional de Estadística - DANE	Equipo técnico

Descripción del elemento:
Ingresar los datos de las personas u organizaciones involucradas en las distintas etapas de la operación estadística: diseño, levantamiento de campo, procesamiento y análisis.

Fuente: DANE.

Financiamiento: relaciona la entidad o las entidades responsables de financiar la operación estadística. Si es posible, se reporta el número de donación o contrato.

Gráfica 26. Financiamiento



The screenshot shows the 'Financiamiento' table with the following data:

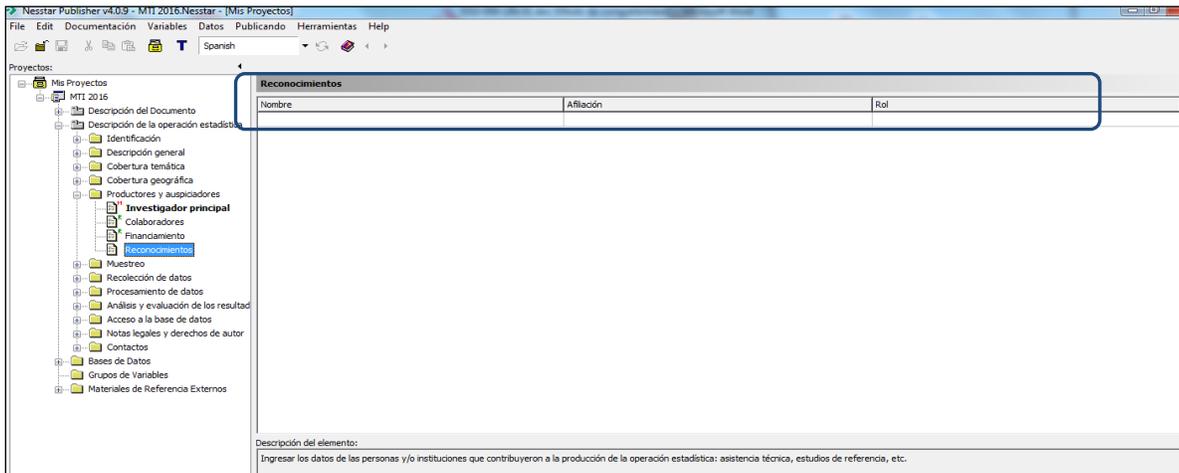
Agencia/Organización	Abreviación	Número de Donación	Rol
Departamento Administrativo Nacional de Estadística	DANE		Editor

Descripción del elemento:
Ingresar los datos de las fuentes de financiamiento de la operación estadística (recursos fiscales y/u otras fuentes). Si existe, especificar el número de donación o contrato.

Fuente: DANE.

Reconocimientos: menciona las personas y/o instituciones que contribuyeron a la producción de la operación estadística.

Gráfica 27. Reconocimientos

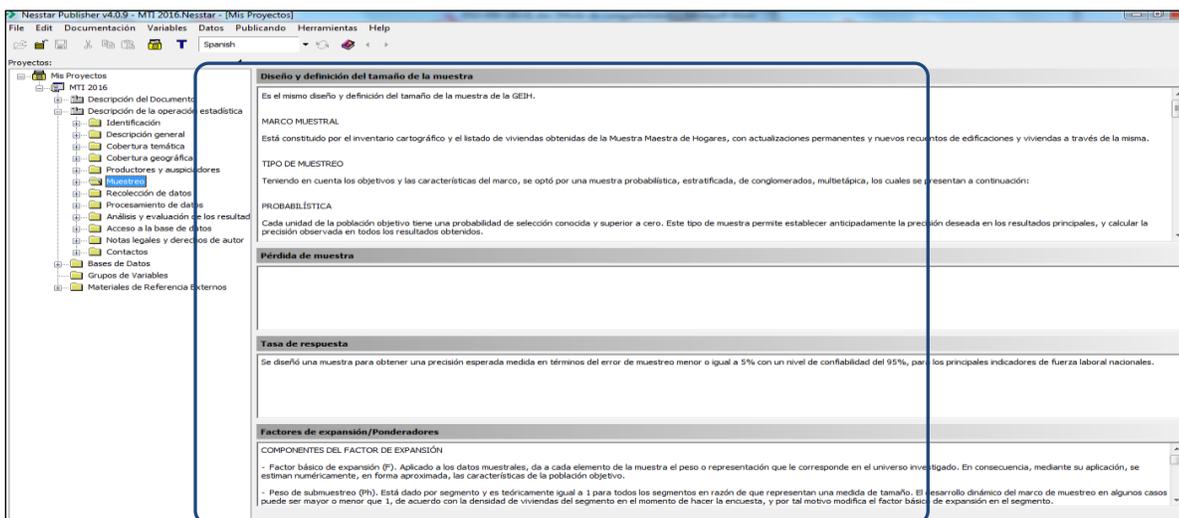


Fuente: DANE.

MUESTREO

En este campo se debe documentar el diseño y la definición del tamaño de la muestra (marco muestral, tipo de muestreo, tamaño final de la muestra), pérdida de muestra, tasa de respuesta y factores de expansión/ponderadores (componente de factor de expansión). Aplica para encuestas por muestreo.

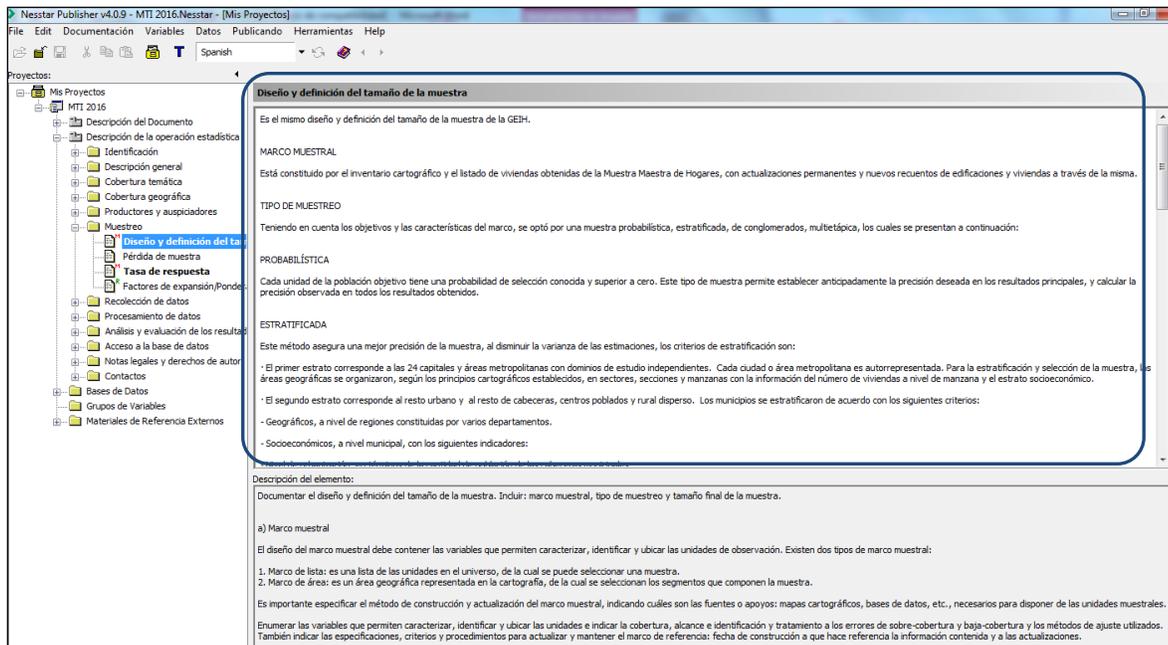
Gráfica 28. Muestreo



Fuente: DANE.

Diseño y definición del tamaño de la muestra: proporciona información sobre el marco muestral, los métodos y los procedimientos utilizados para seleccionar a los encuestados y el tamaño de muestra. Los títulos de cada uno de los temas deben quedar en mayúscula.

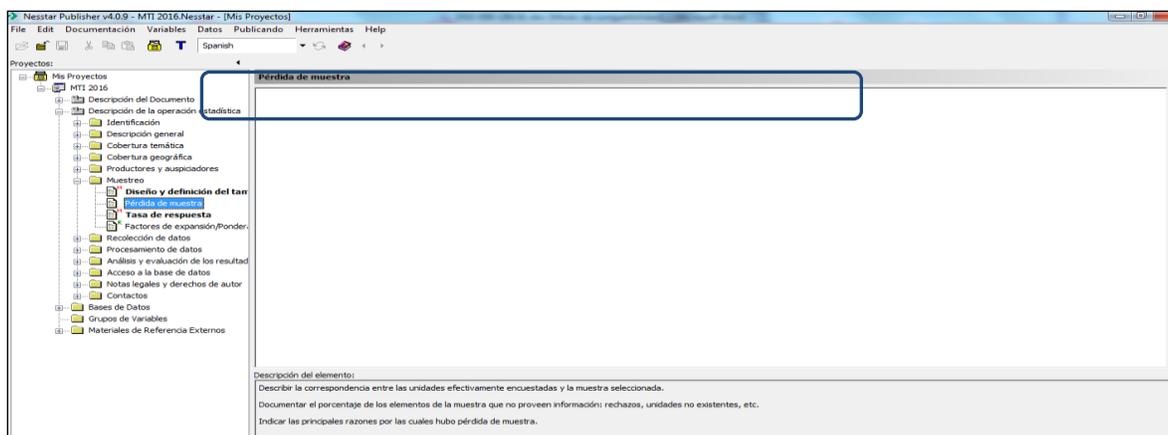
Gráfica 29. Diseño y definición del tamaño de la muestra



Fuente: DANE.

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

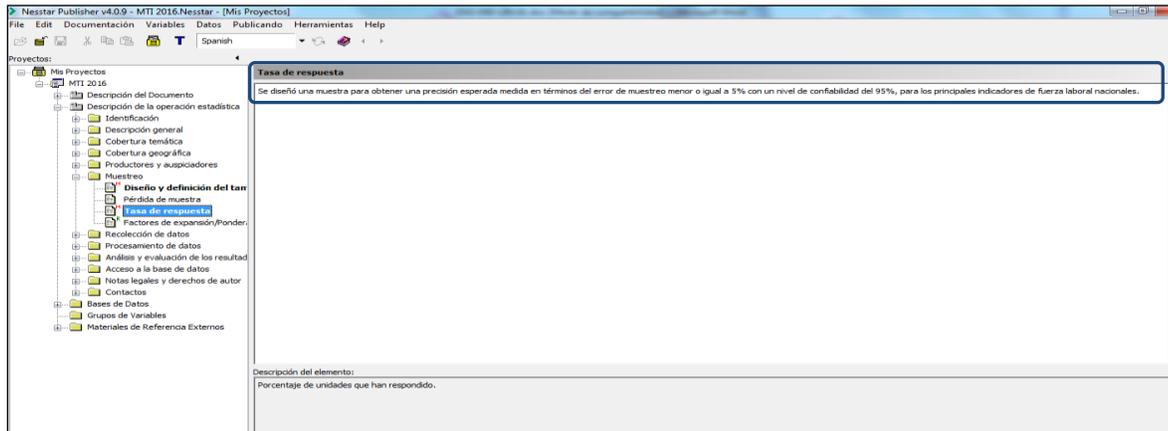
Gráfica 30. Pérdida de muestra



Fuente: DANE.

Tasa de respuesta: se diligencia el porcentaje de respuesta de las unidades seleccionadas.

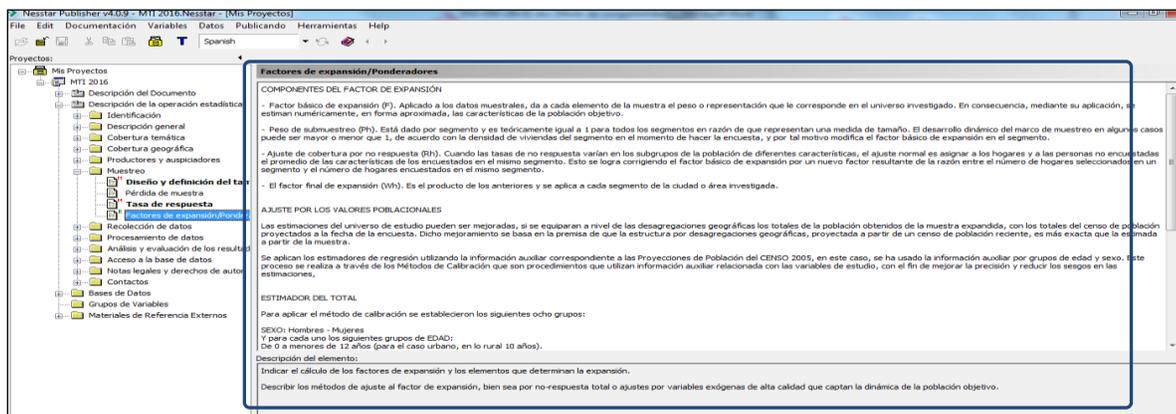
Gráfica 31. Tasa de respuesta



Fuente: DANE.

Factores de expansión/ponderadores: relaciona el cálculo de los factores de expansión y los elementos que determinan la expansión, es necesario describir los métodos de ajuste al factor de expansión, bien sea por no respuesta total o ajustes por variables exógenas de alta calidad que captan la dinámica de la población objetivo.

Gráfica 32. Factores de expansión/ponderadores



Fuente: DANE.

RECOLECCIÓN DE DATOS

En este campo se relaciona la siguiente información: periodo de recolección, periodo de referencia, método de recolección de los datos, notas sobre la recolección de los datos, supervisión y control del operativo de recolección y recolector de datos.

Se recomienda para este campo incluir un informe completo sobre las actividades de recolección de datos y una copia de todos los cuestionarios utilizados durante su ejecución.

Gráfica 33. Recolección de datos

The screenshot shows the 'Recolección de datos' section in the Neustar Publisher v4.0.9 software. The interface is divided into several sections:

- Periodo de recolección:** A table with columns 'Inicio', 'Fin', and 'Ciclo'. The 'Inicio' field contains '2016' and the 'Fin' field contains '2016'. The 'Ciclo' field is empty.
- Periodo de referencia:** A table with columns 'Inicio', 'Fin', and 'Ciclo'. The 'Inicio' field contains '2016' and the 'Fin' field contains '2016'. The 'Ciclo' field contains 'Cuarto Trimestre'.
- Método de recolección de los datos:** A dropdown menu with the selected option 'Entrevista personal asistida con PDA (dispositivo móvil de captura)'.
- Notas sobre la recolección de los datos:** A text area containing the following text:

SENSIBILIZACIÓN
El Sistema de Sensibilización, cuyo objetivo principal es: "Lograr una respuesta oportuna y de calidad por parte de las fuentes de información para la investigación estadística al DANE", busca lograr un ambiente de colaboración, acercamiento y conocimiento de la población objetivo mediante la visita a cada uno de los hogares a los que les será aplicada la encuesta, por parte de los sensibilizadores, con el fin de informarles que el hogar ha sido seleccionado para ser encuestado. En esta primera visita, se pueden establecer citas y horarios en los cuales los encuestadores del DANE pueden aplicar la encuesta.
- Supervisión y control del operativo de recolección:** A text area containing the following text:

La recolección de la información está bajo la dirección del supervisor, quien fundamentalmente controla el trabajo de campo que realizan los encuestadores, y cuyos resultados dependen en gran parte de la iniciativa para la solución de problemas y dificultades, del liderazgo y la capacidad organizativa del trabajo.
Las actividades de control que debe llevar a cabo un supervisor son:
- Control de desempeño de las actividades diarias del encuestador.
- Recolector de datos:** A table with columns 'Nombre', 'Abreviación', and 'Afilación'. The 'Nombre' field contains 'Departamento Administrativo Nacional de Estadística', the 'Abreviación' field contains 'DANE', and the 'Afilación' field contains 'Gobierno Nacional'.

Fuente: DANE.

Periodo de recolección: relaciona las fechas de inicio y fin de la recolección de datos indicando el ciclo de este (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.

Gráfica 34. Periodo de recolección

Inicio	Fin	Ciclo
2016	2016	

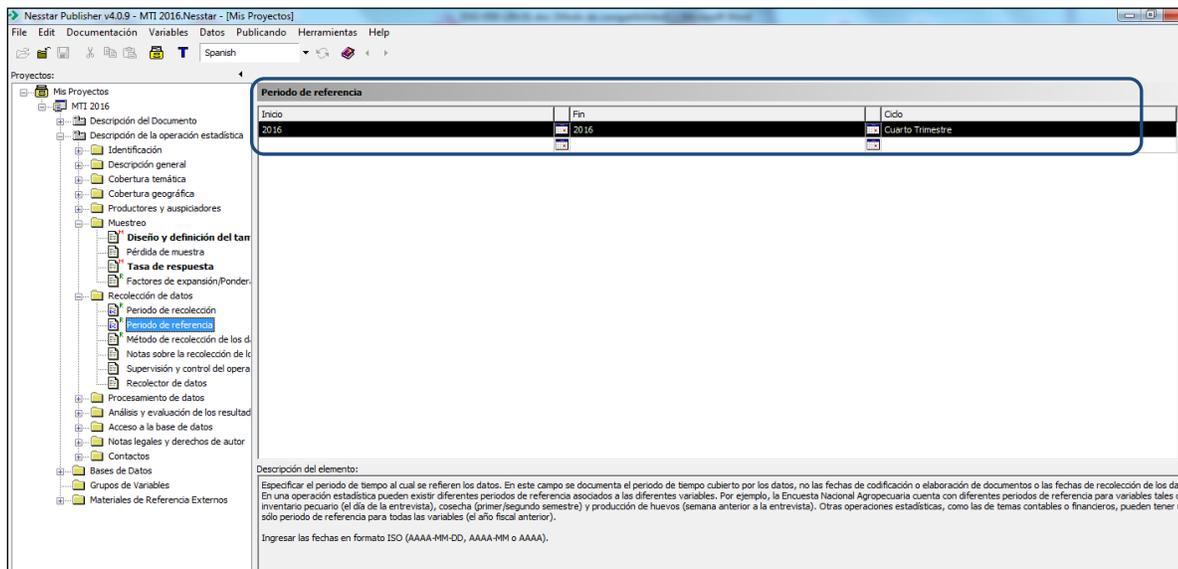
Descripción del elemento:
 Ingresar las fechas en que se realiza la recolección de los datos e indicar el ciclo (mensual, trimestral, anual, etc.).
 Ingresar las fechas en formato ISO (AAAA-MM-DD, AAAA-MM o AAAA).

Fuente: DANE.

Periodo de referencia: debe especificar el periodo de tiempo al que se refieren los datos, no las fechas de codificación o de elaboración de documentos o las fechas de recolección de los datos.

En una operación estadística pueden existir diferentes periodos de referencia asociados a las diferentes variables. Por ejemplo, la Encuesta Nacional Agropecuaria cuenta con diferentes periodos de referencia para variables tales como inventario pecuario (el día de la entrevista), cosecha (primer/segundo semestre) y producción de huevos (semana anterior a la entrevista). Otras operaciones estadísticas, como por ejemplo las Estadísticas de Monitoreo y Seguimiento de la Calidad del Aire del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) pueden tener un único periodo de referencia para todas las variables (el mes inmediatamente anterior al cargue de datos por parte de los operadores del Sistema de Vigilancia de la Calidad del Aire (SVCA) que reportan información al Subsistema de Información sobre Calidad del Aire (SISAIRE)).

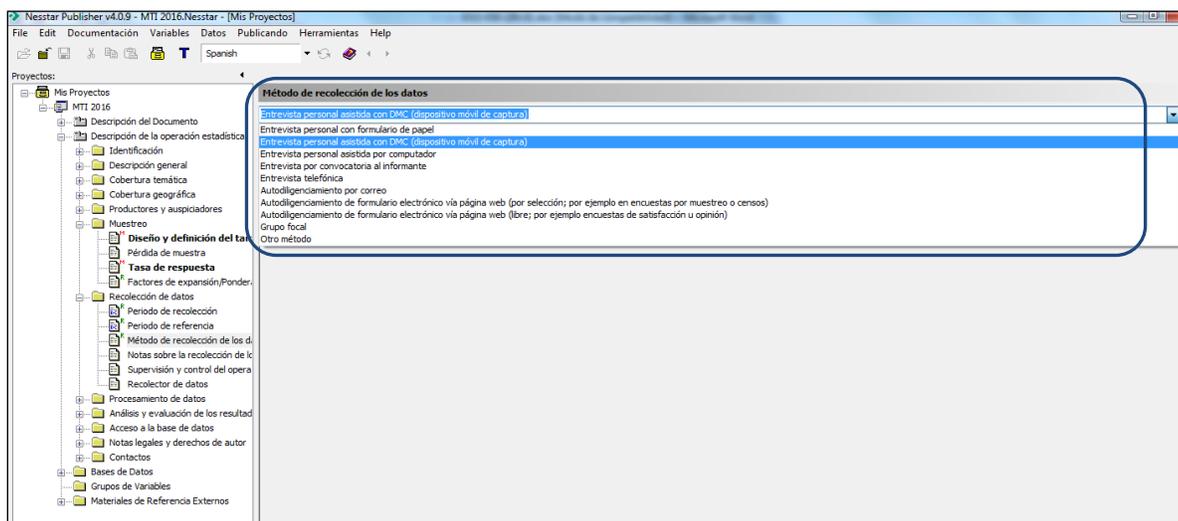
Gráfica 35. Periodo de referencia



Fuente: DANE.

Método de recolección: se selecciona de la lista desplegable la opción que aplique a la operación estadística.

Gráfica 36. Método de recolección

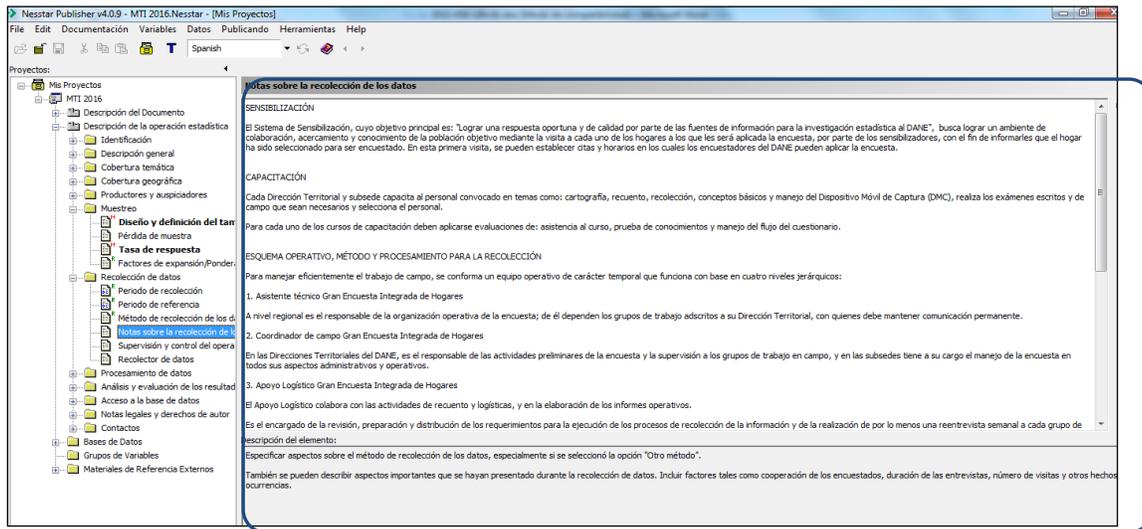


Fuente: DANE.

Notas sobre la recolección de datos: se debe especificar los 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 relacionados con la recolección.

Gráfica 37. Notas sobre la recolección de los datos

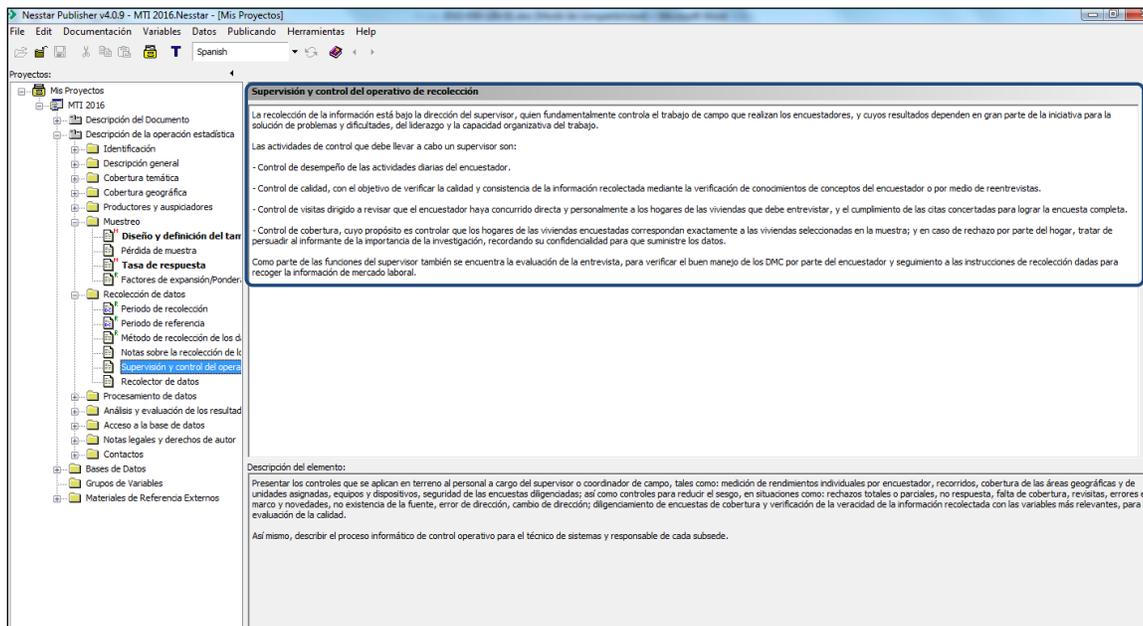


Fuente: DANE.

Supervisión y control del operativo de recolección: describe los controles que se aplican en terreno al personal a cargo de la recolección, la supervisión o la coordinación de campo, tales como: recorridos, cobertura de las áreas geográficas y de unidades asignadas, equipos y dispositivos, seguridad de los cuestionarios diligenciados. También se incluye los controles para reducir el sesgo, en situaciones como: rechazos totales o parciales; no respuesta; falta de cobertura; visitas; 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.

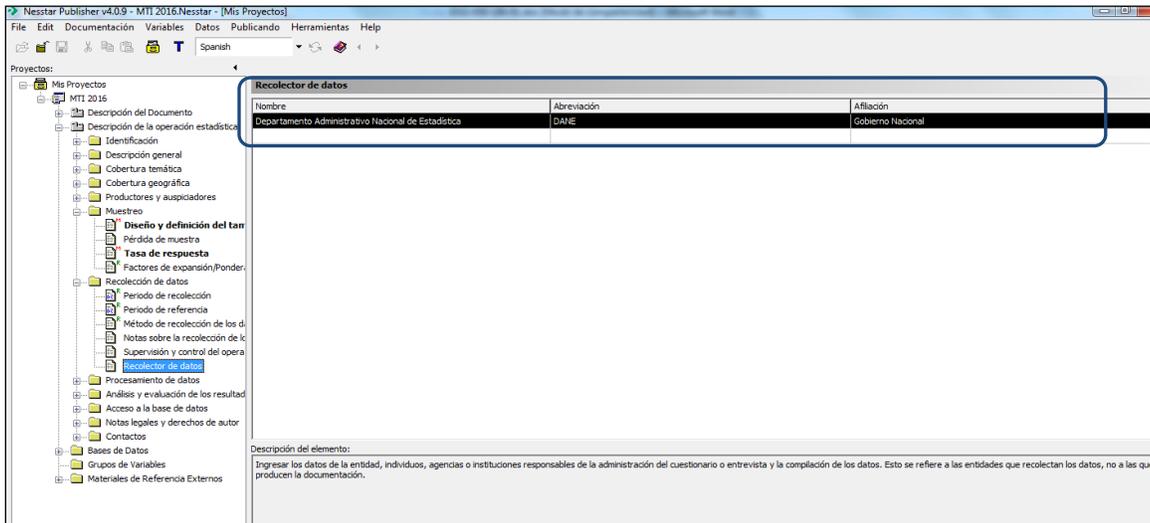
Asimismo, debe incluir un resumen de los procedimientos seguidos para minimizar errores, reducir pérdida de información y prevenir falsificación. Estos procedimientos pueden involucrar revisiones de los supervisores de los cuestionarios diligenciados en el campo, re-entrevistas a los encuestados y otras actividades similares.

Gráfica 38. Supervisión y control del operativo de recolección



Fuente: DANE.

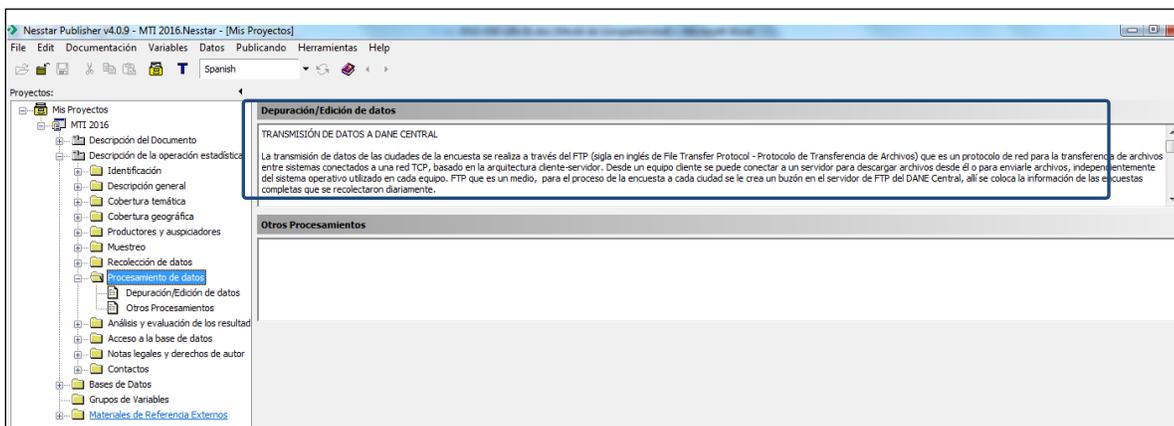
Recolector de datos: relaciona el listado de las personas y las organizaciones responsables de administrar los cuestionarios y recolección de datos. Al momento de diligenciar se debe tener en cuenta que las primeras letras del nombre de la entidad, la abreviación de la entidad y la afiliación deben ir en mayúscula. Para entidades públicas del orden nacional la afiliación corresponde a Gobierno Nacional, para el caso de las secretarías departamentales la afiliación será departamental y para el caso de los gremios la afiliación será gremial.

Gráfica 39. Recolector de datos

Fuente: DANE.

PROCESAMIENTO DE DATOS

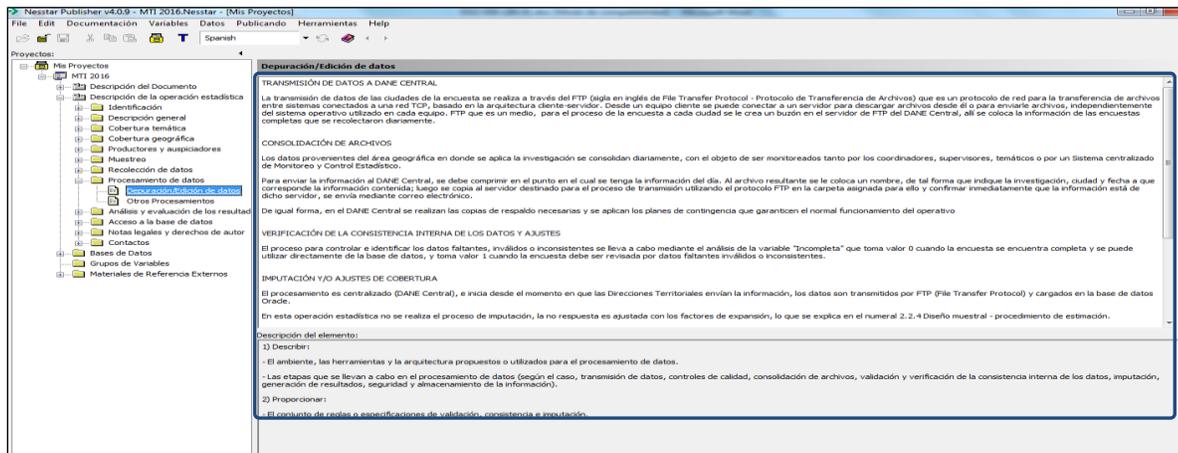
Está conformado por un grupo de campos donde se relaciona las técnicas aplicadas para llevar a cabo la depuración/edición de datos y otros procedimientos como la integración de datos, la clasificación, la codificación, la derivación de nuevas variables y unidades, la revisión y la validación de los datos, el cálculo de ponderadores, la generación de resultados y la evaluación de los archivos de datos.

Gráfica 40. Procesamiento de datos

Fuente: DANE.

Depuración/edición de datos: describe el ambiente, las herramientas y la arquitectura propuestas o utilizadas para el procesamiento de dato, las etapas que se llevan a cabo en el procesamiento de datos (según el caso, transmisión de datos, controles de calidad, consolidación de archivos, validación y verificación de la consistencia interna de los datos, imputación, generación de resultados, seguridad y almacenamiento de la información), además relaciona el conjunto de reglas o las especificaciones de validación, consistencia e imputación.

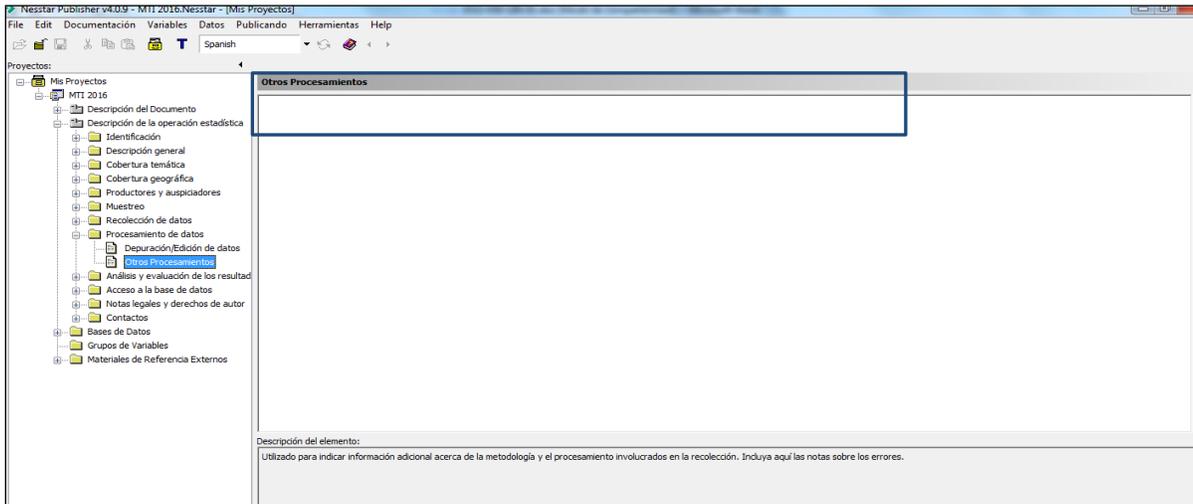
Gráfica 41. Depuración/edición de datos



Fuente: DANE.

Otros procesamientos: se consigna la información adicional acerca de la metodología y el procesamiento involucrados en la recolección y las notas relacionada con los errores.

Gráfica 42. Otros procesamientos

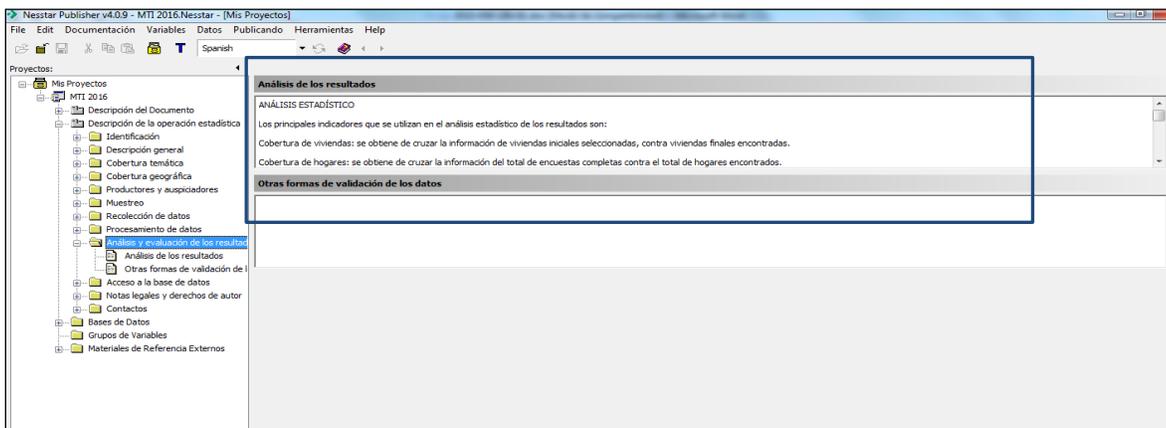


Fuente: DANE.

ANÁLISIS Y EVALUACIÓN DE LOS RESULTADOS

En este campo se relaciona información relacionada con los análisis de los resultados y otras formas de validación de los datos.

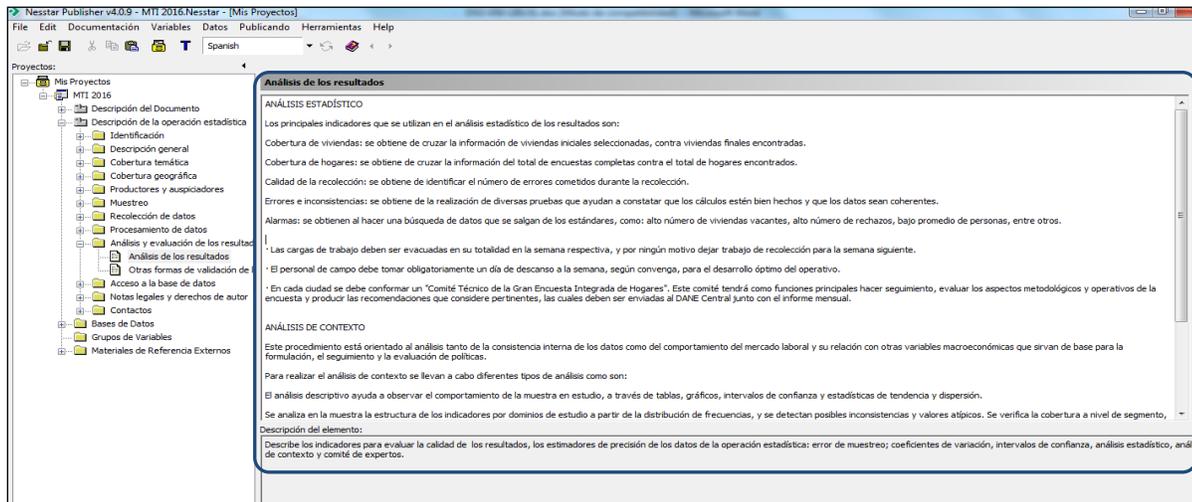
Gráfica 43. Análisis y evaluación de los resultados



Fuente: DANE.

Análisis de los resultados: describe los indicadores para evaluar la calidad de los resultados, los estimadores de precisión de los datos de la operación estadística (error de muestreo, coeficientes de variación e intervalos de confianza) así como el análisis estadístico, el análisis de contexto y los comités de expertos que se realicen para revisar los resultados de la operación estadística.

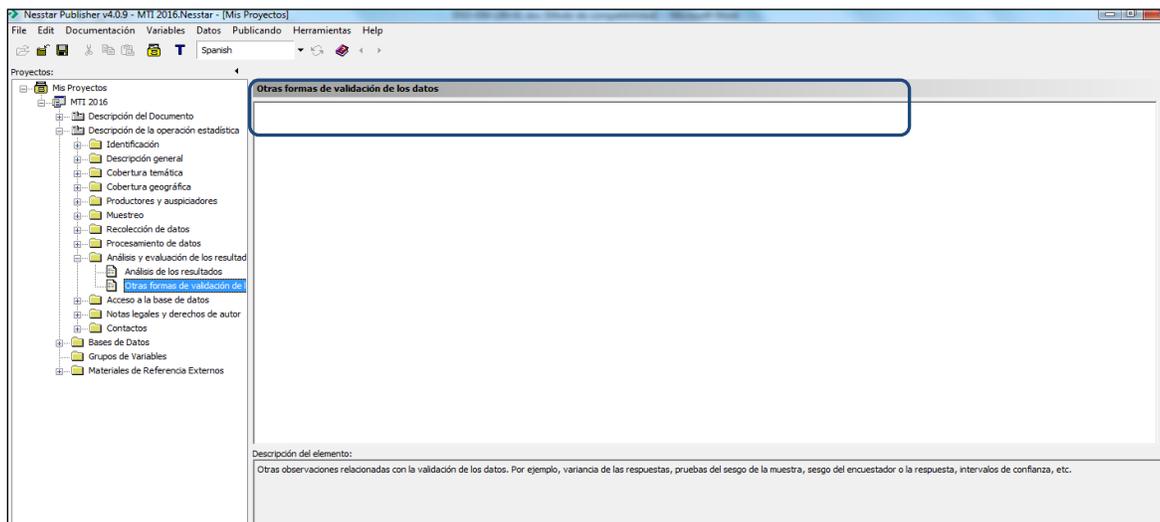
Gráfica 44. Análisis de los resultados



Fuente: DANE.

Otras formas de validación de los datos: incluye las observaciones relacionadas con la validación de los datos. Por ejemplo: variancia de las respuestas, pruebas del sesgo de la muestra, sesgo del encuestador o la respuesta, intervalos de confianza, etc.

Gráfica 45. Otras formas de campo de validación de los datos

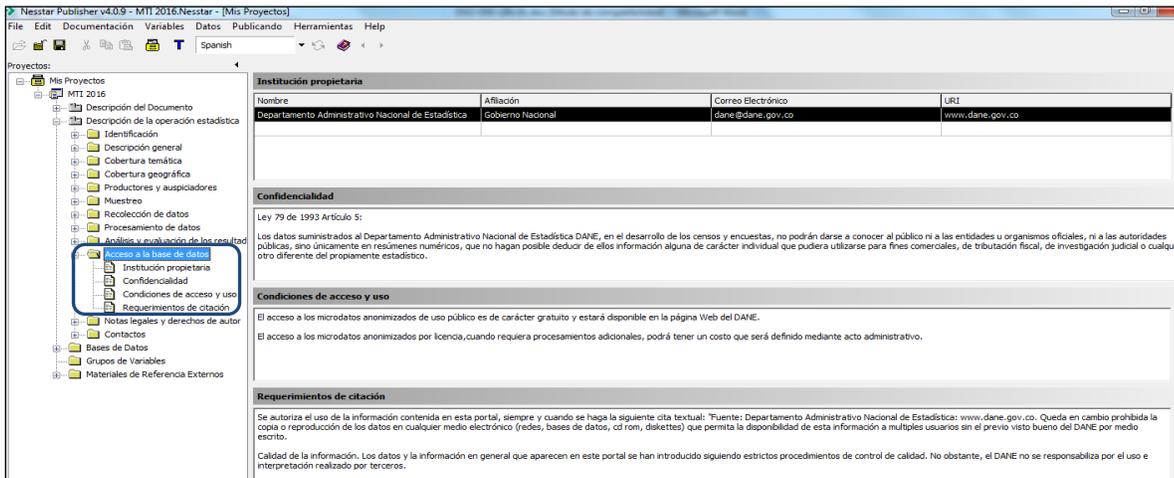


Fuente: DANE.

ACCESO A LAS BASES DE DATOS

En este campo se relaciona información respecto a la institución propietaria; confidencialidad; condiciones de acceso y uso, y requerimientos de citación.

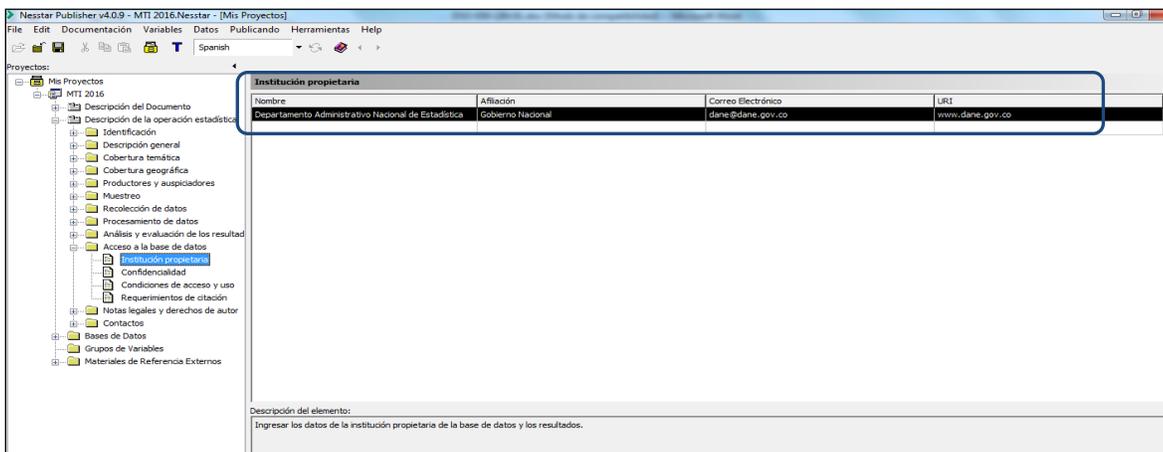
Gráfica 46. Acceso a las bases de datos



Fuente: DANE.

Institución propietaria: debe relacionar los nombres de las personas y las organizaciones responsables de otorgar acceso a los datos y la documentación.

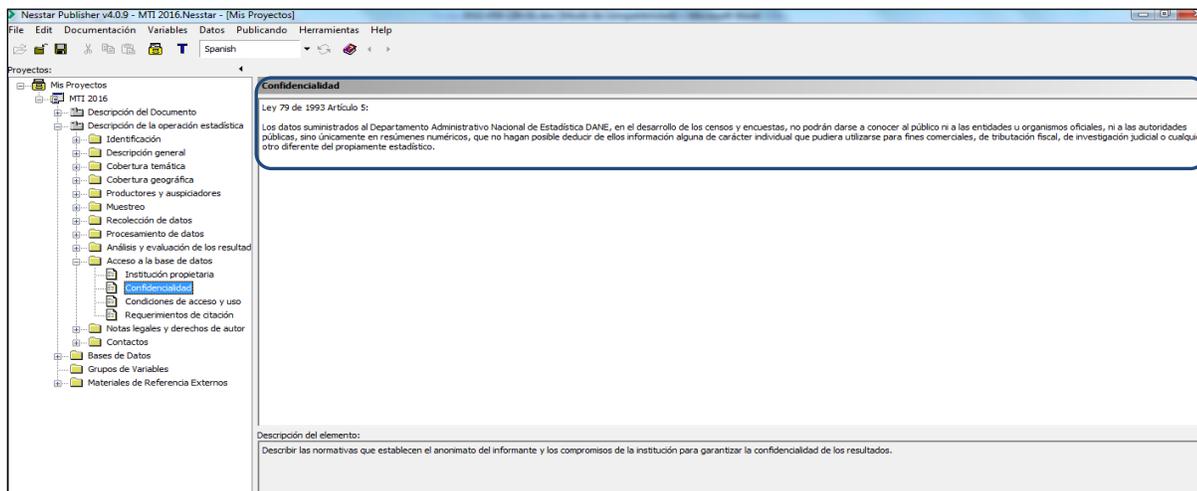
Gráfica 47. Institución propietaria



Fuente: DANE.

Confidencialidad: describe la normatividad que establece el secreto estadístico de los informantes y los compromisos de la institución para garantizar la confidencialidad de los resultados.

Gráfica 48. Confidencialidad

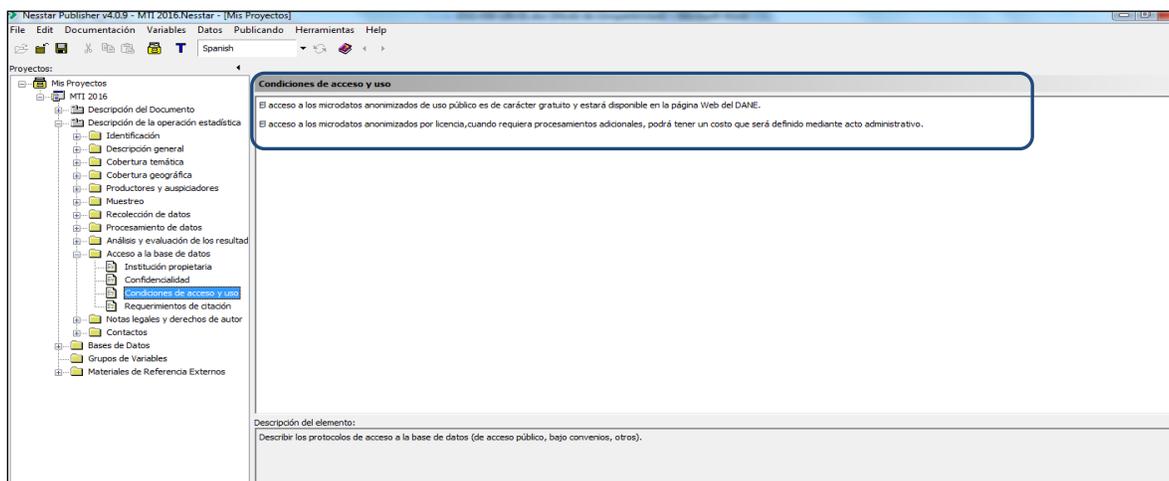


Fuente: DANE.

Condiciones de acceso y uso: describe 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.

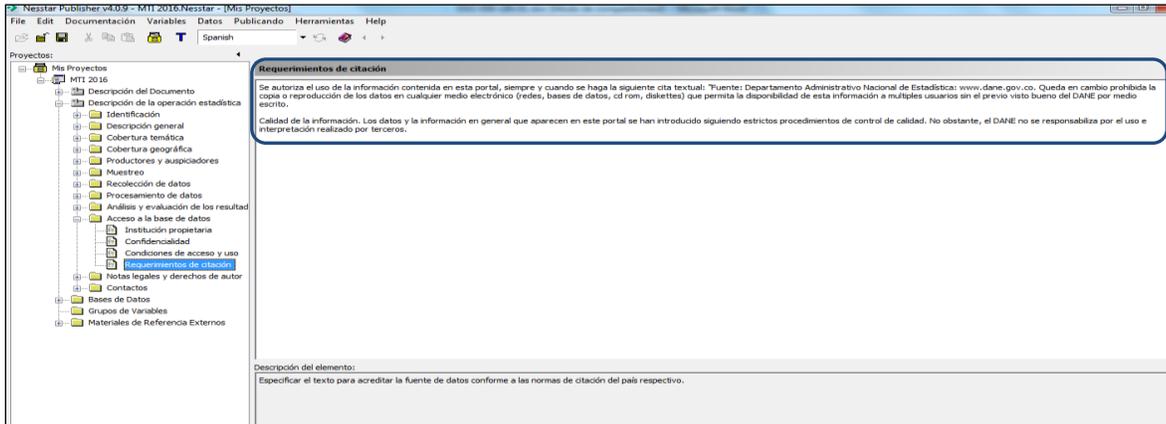
Gráfica 49. Condiciones de acceso y uso



Fuente: DANE.

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

Gráfica 50. Requerimientos de citación

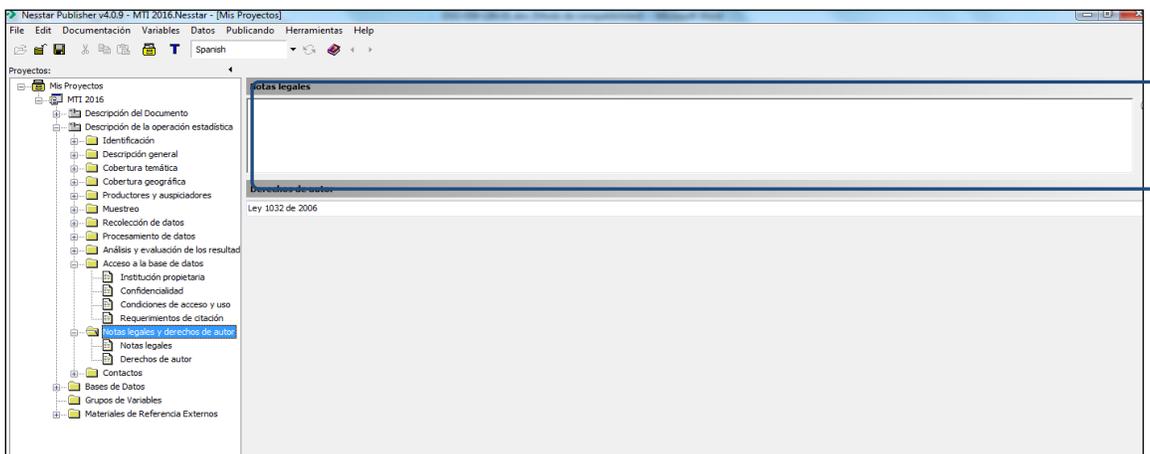


Fuente: DANE.

NOTAS LEGALES Y DERECHOS DE AUTOR

En este campo se provee información relativa a la responsabilidad de los usuarios de la documentación y las bases de datos: notas legales y derechos de autor.

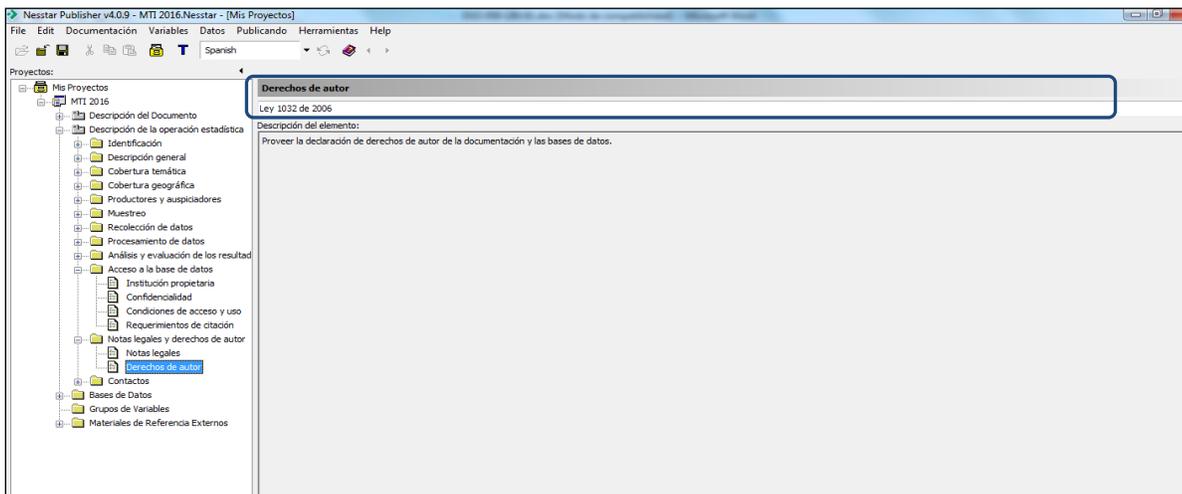
Gráfica 51. Notas legales



Fuente: DANE.

Derechos de autor: documenta la declaración de derechos de autor de los resultados y debe: especificar la autoría y la propiedad intelectual y otorgar protección a las fuentes de información y a la entidad generadora de los indicadores.

Gráfica 52. Derechos de autor

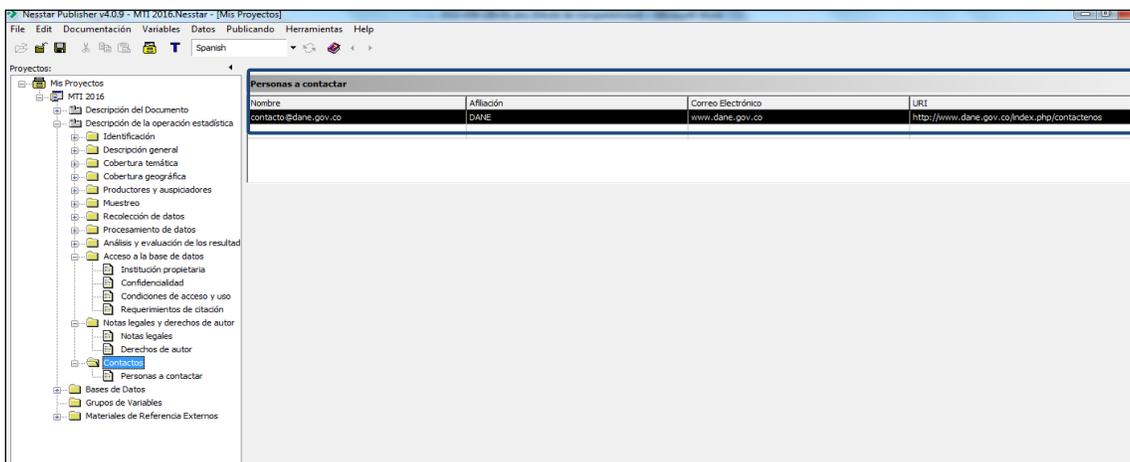


Fuente: DANE.

CONTACTOS

Este campo se utiliza para proporcionar a los usuarios información acerca de quién puede responder preguntas adicionales sobre la operación estadística: personas a contactar.

Gráfica 53. Personas a contactar



Fuente: DANE.

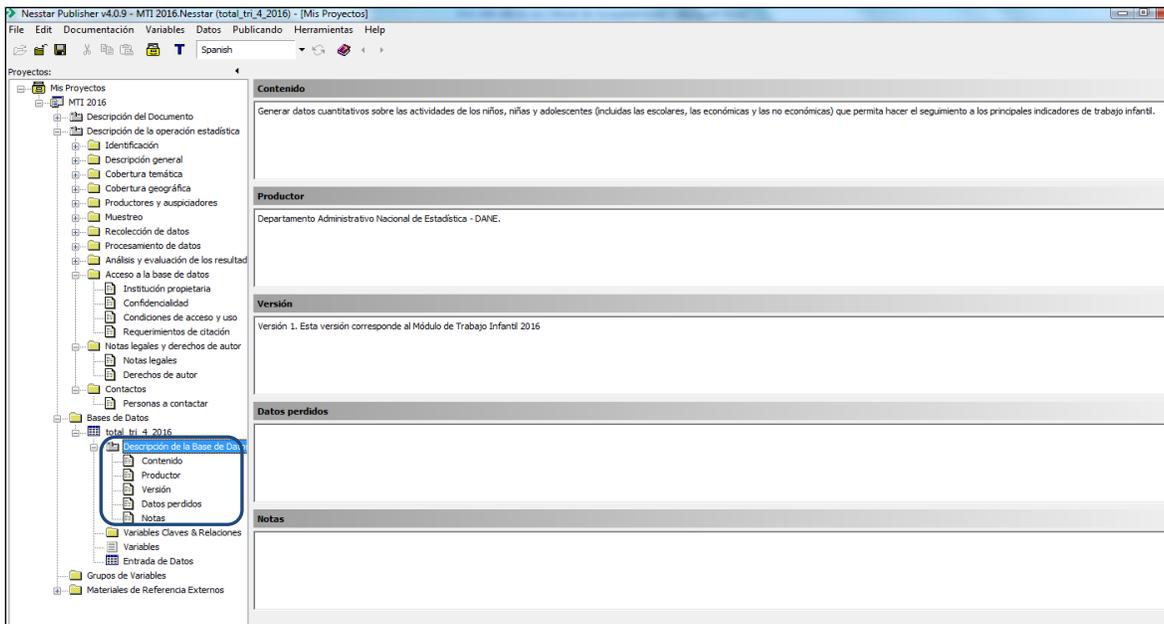
2.3 Base de datos

Esta sección contiene todos los archivos de datos asociados con la operación estadística, incluye una descripción detallada de cada variable. Para cada archivo de datos se debe documentar cuatro campos, a saber: descripción de la base de datos, variables clave y relaciones, variables y entrada de datos.

DESCRIPCIÓN DE LA BASE DE DATOS

Este campo se proporciona información básica acerca de cada archivo como el contenido, el productor, la versión y los datos perdidos y notas.

Gráfica 54. Descripción de la base de datos



Fuente: DANE.

Contenido: describe las bases de datos del archivo de datos seleccionado incluyendo cobertura temática, características especiales del contenido y principales variables.

Productor: documenta el nombre de la persona o la 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 entidad que recolectó o procesó los datos.

Versión: describe las versiones del archivo de datos, contiene información de fecha, origen y tipo de datos (validados, parcialmente validados, originales, etc.).

Es recomendable asignar un número de versión a las diferentes publicaciones de la base de dato, al documentar se debe poner la palabra «versión» y el año de la operación estadística, como por ejemplo versión 2005.

Datos perdidos: se usa para describir las causas por las que existen datos perdidos en el archivo de datos. Ejemplo: fusión de archivos, error de programación, etc.

Notas: se puede ingresar comentarios adicionales acerca del archivo de datos.

VARIABLES CLAVES Y RELACIONES

La mayoría de los estudios incluyen múltiples archivos de datos relacionados. Este grupo de campos están diseñados con el propósito de definir las relaciones entre estos archivos.

Dos tipos de archivos son utilizados 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, ya que las «variables clave base» serán utilizadas por los analistas de datos para fusionar los archivos en STATA7, SPSS.SAV, DTA o cualquier otro programa estadístico.

VARIABLES

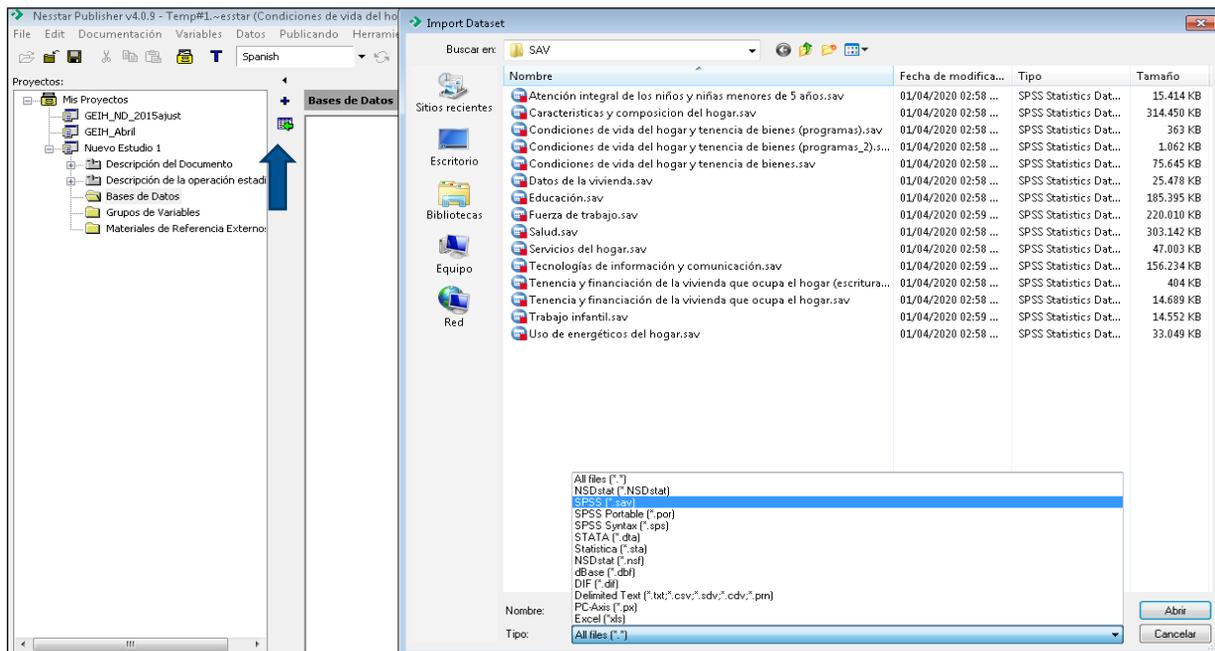
Este campo se usa para ingresar los metadatos de cada variable, comprende el listado de variables y su etiqueta o descripción. Después del cargue de las variables de la base de datos se inicia la documentación.

Es importante tener claro qué se pueden documentar dos tipos de información: bases de datos de operaciones estadísticas con microdatos y documentación de la estructura de la base de datos (Diccionario de datos)

Documentación de base de datos con microdatos

- Con el icono Importar la base de datos se selecciona el formato con microdatos (ejemplo: txt, stata, spss, dbf, entre otros).
- Se verifica que los microdatos de la base de datos coincidan con la estructura que va a aparecer en el aplicativo.
- Es importante recordar que debo tener la información básica como el código y la descripción de las variables para poder identificarlas cuando se carga al sistema.

Gráfica 55. Importar base de datos



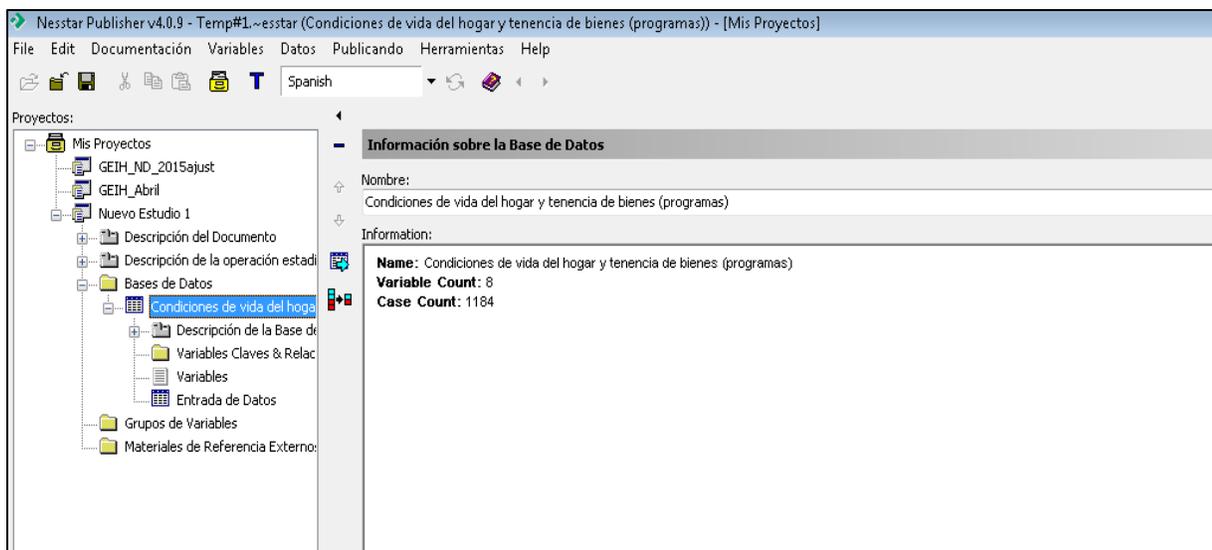
Fuente: DANE.

Después de haber importado la base de datos se despliegan cuatro campos diferentes para documentar:

1. Descripción de la base de datos: los cuales son elementos proporcionados para saber la información básica acerca de las bases de datos.

2. Las variables Claves & Relaciones: sirven para definir relaciones de variables entre varias bases de datos.
3. Variables: en este campo se observa la estructura de base de datos para documentar los metadatos.
4. Entrada de datos: En este campo se visualiza la base de datos importada.

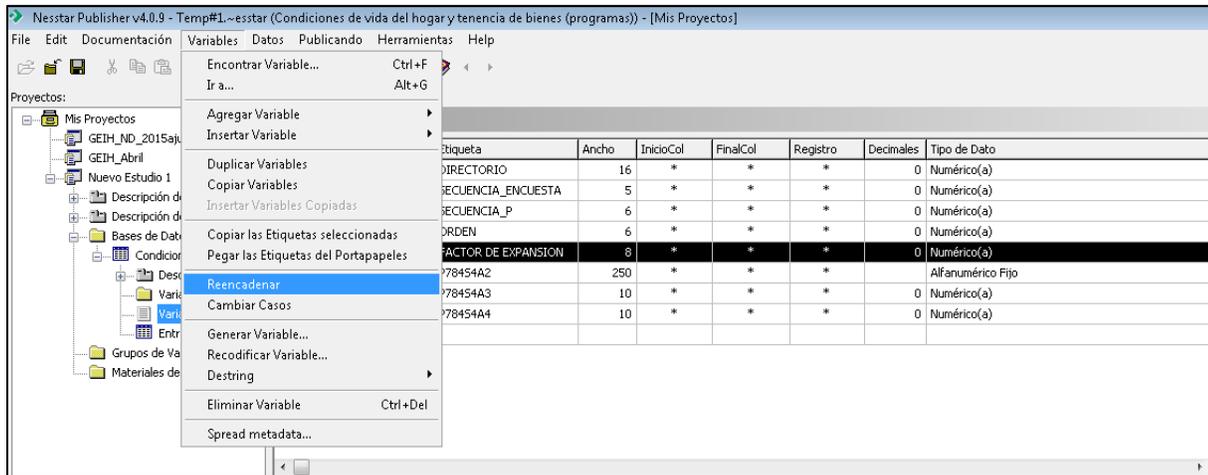
Gráfica 56. Campos para documentar



Fuente: DANE.

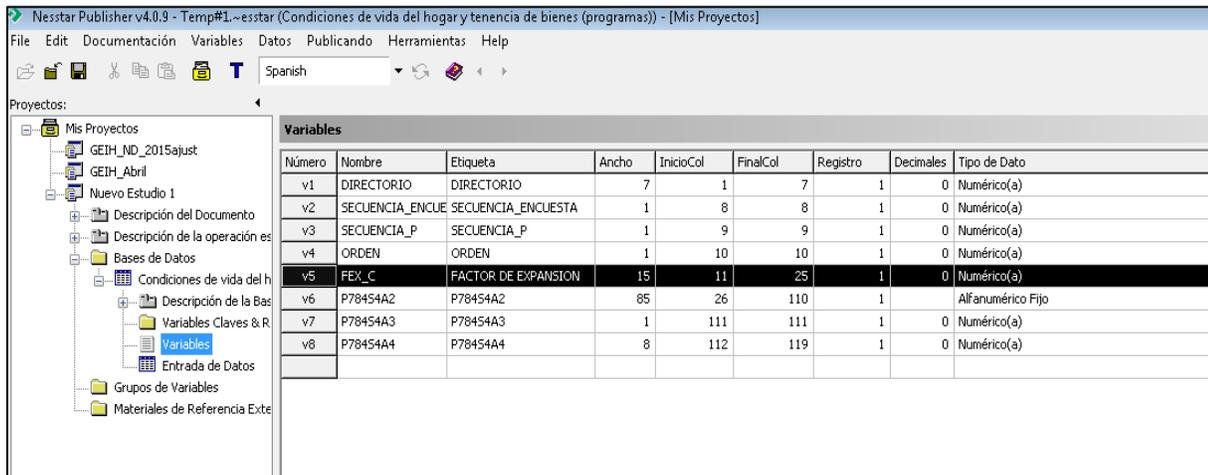
Después ir al menú y en la pestaña variables se selecciona **reencadenar** cuyo objetivo es cargar los datos del ancho, inicio y final de la columna. Automáticamente carga la información de las variables y sus características, es decir: (numérico, alfanumérico y/o fecha).

Gráfica 57. Reencadenar



Fuente: DANE.

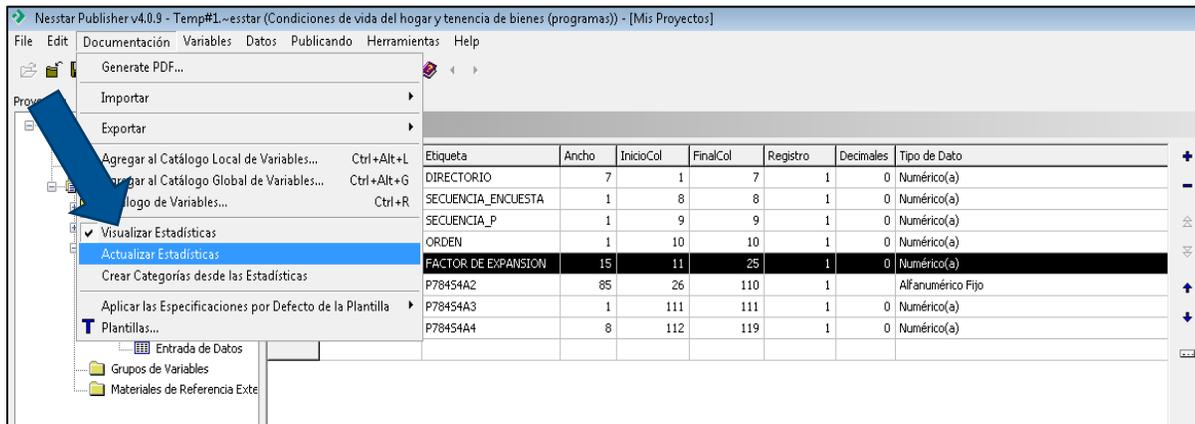
Gráfica 58. Carga de Ancho, InicioCol, FinalCol y Tipo de Dato



Fuente: DANE.

Luego se ubica la pestaña **documentación** y se selecciona **actualizar estadísticas**.

Gráfica 59. Actualización de estadísticas



Fuente: DANE.

En la parte de documentación se tienen tres pestañas:

1. Estadísticas: el software 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.
2. 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.
3. 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.

Gráfica 60. Sección documentación

Nesstar Publisher v4.0.9 - Temp#1.-.esstar (Condiciones de vida del hogar y tenencia de bienes (programas)) - [Mis Proyectos]

File Edit Documentación Variables Datos Publicando Herramientas Help

Spanish

Proyectos:

- Mis Proyectos
 - GEIH_MD_2015ajust
 - GEIH_Abril
 - Nuevo Estudio 1
 - Descripción del Documento
 - Descripción de la operación es
 - Bases de Datos
 - Condiciones de vida del h
 - Descripción de la Bas
 - Variables Claves & R
 - Variables
 - Entrada de Datos
 - Grupos de Variables
 - Materiales de Referencia Exte

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v1	DIRECTORIO	DIRECTORIO	7	1	7	1	0	N Numérico(a)
v2	SECUENCIA_ENCUE	SECUENCIA_ENCUESTA	1	8	8	1	0	N Numérico(a)
v3	SECUENCIA_P	SECUENCIA_P	1	9	9	1	0	N Numérico(a)
v4	ORDEN	ORDEN	1	10	10	1	0	N Numérico(a)
v5	FEX_C	FACTOR DE EXPANSION	15	11	25	1	0	N Numérico(a)
v6	P78454A2	P78454A2	85	26	110	1		Alfanumérico Fijo
v7	P78454A3	P78454A3	1	111	111	1	0	N Numérico(a)
v8	P78454A4	P78454A4	8	112	119	1	0	N Numérico(a)

Documentación

Estadísticas Ponderadores Documentación

Fuente: DANE.

El software Nesstar Publisher proporciona la pestaña **documentación** y presenta cuatro elementos: descripción, preguntas, imputación y derivación y otros.

Gráfica 61. Elementos de la documentación

The screenshot shows the Nesisstar Publisher v4.0.9 interface. On the left is a project tree with folders like 'Método de recolección', 'Procesamiento de datos', and 'Bases de Datos'. The center panel displays a 'Variables' table with columns: Número, Nombre, Etiqueta, Ancho, InicioCol, FinalCol, Registro, Decimales, and Tipo de Dato. The right panel shows 'Documentación' with sub-sections 'Estadísticas' and 'Ponderadores'. Under 'Estadísticas', there is a tree structure including 'Definición', 'Universo', 'Encuestado idóneo', 'Preguntas', 'Imputación y derivación', and 'Otros'. A blue arrow points to the 'Definición' and 'Universo' items. Below the tree is a text area for 'Descripción del elemento:' with the instruction: 'Describir de forma detallada la variable, proporcionando información más completa que aquella provista por el en su etiqueta.'

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v1	CIUDAD	CIUDAD	2	1	2	1	0	N Numérico(a)
v2	CODPTE	Código de la fuente	5	3	7	1	0	Alfanumérico Fijo
v3	ANOPRE	Año del precio	4	8	11	1	0	N Numérico(a)
v4	MESPRE	Mes del precio	2	12	13	1	0	N Numérico(a)
v5	CODART	Código del artículo	4	14	17	1	0	Alfanumérico Fijo
v6	INV	Investigación	3	18	20	1	0	N Numérico(a)
v7	Nombre	Nombre del artículo	5	21	25	1	0	Alfanumérico Fijo
v8	Unidadbase	Unidad base que se debe re	14	26	39	1	0	N Numérico(a)
v9	Unidadanterior	Unidad anterior recolectada	2	40	41	1	0	N Numérico(a)
v10	PREANT	Precio Base anterior	2	42	43	1	0	Alfanumérico Fijo

Fuente: DANE.

En **Descripción** se deben diligenciar los siguientes campos:

- **Definición:** permite describir con mayor detalle la variable.
- **Universo:** permite especificar la población a la que aplica la variable.
- **Encuestado idóneo:** documenta la información de 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 62. Descripción

The screenshot shows the Nesstar Publisher v4.0.9 interface. The 'Variables' table is visible, listing variables v1 through v8. The 'Definición' section for variable v1 is expanded, showing the 'Universo' and 'Encuestado idóneo' sections.

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v1	DIRECTORIO	DIRECTORIO	7	1	7	1	0	Número(a)
v2	SECUENCIA_P	SECUENCIA_P	1	8	8	1	0	Número(a)
v3	TAM_HOGAR	Tamaño del hogar	2	9	10	1	0	Número(a)
v4	ORDEN	ORDEN	2	11	12	1	0	Número(a)
v5	RAMAZO	¿A qué actividad se dedica principalmente la empres	2	13	14	1		Alfanumérico l
v6	RAMAZO	¿A qué actividad se dedica principalmente la empres	4	15	18	1		Alfanumérico l
v7	CLASE	Clase(estrato/cobertura)	1	19	19	1		Alfanumérico l
v8	OFICIO	¿Qué hace . . . en este trabajo? Codificada	2	20	21	1	0	Número(a)

The 'Definición' section for variable v1 shows the following text:

Universo
Para el caso del Módulo de Trabajo Infantil, el universo está conformado por la población de 5 a 17 años de edad civil no institucional residente en viviendas particulares.

Encuestado idóneo
Directo

Fuente: DANE.

Se despliega la carpeta **Preguntas** y se diligencia cada uno de los campos de la siguiente manera:

- El elemento «Texto» de la pre-pregunta puede ser utilizado para describir la pregunta inmediatamente anterior, según las condiciones del formulario.
- Se debe ingresar 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.

Gráfica 63. Preguntas

The screenshot shows the Nesstar Publisher v4.0.9 interface. The main window displays a table of variables and a detailed documentation view for a selected variable.

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v40	P42257	g. Limpieza y mantenimiento del hogar.	1	62	62	1	0	Número(a)
v41	P42258	h. Otros	1	63	63	1	0	Número(a)
v42	P423	¿Por qué razón principal <...> realiza o colabora en	1	64	64	1	0	Número(a)
v43	P405	¿Por qué razón principal trabaja <...>?	1	65	65	1	0	Número(a)
v44	P408	En este trabajo <...> es:	1	66	66	1	0	Número(a)
v45	P409	Antes de descuentos ¿cuánto ganó ...el mes pasado	7	67	73	1	0	Número(a)
v46	P410	Además del salario en dinero, ¿el mes pasado recibí	1	74	74	1	0	Número(a)
v47	P41051	¿En cuánto estima lo que recibió?	6	75	80	1	0	Número(a)

The documentation view for variable v44 (P408) shows the following structure:

- Pre-pregunta:** 3. ¿Por qué razón principal trabaja <...>?
 - 1- Debe costearse el estudio
 - 2- Debe ayudarse con los gastos de la casa
 - 3- Debe participar en la actividad económica de la familia
 - 4- El trabajo lo forma y lo hace honrado
 - 5- El trabajo lo aleja de los vicios
- Pregunta textual:** 4. En este trabajo <...> es:
 - 1 - Obrero o empleado de empresa particular
 - 2 - Obrero o empleado del gobierno
 - 3 - Empleado doméstico
 - 4 - Trabajador por cuenta propia
 - 5 - Patrón o empleador
- Post-pregunta:** 5. Antes de descuentos ¿cuánto ganó ...el mes pasado en este empleo?
- Manual del entrevistador:** (Empty field)

Fuente: DANE.

En el campo **imputación y derivación** se documentan las imputaciones y las variables derivadas o generadas, lo cual facilita a los usuarios la reproducción de 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.

Gráfica 64. Imputación y derivación

The screenshot shows the Nesstar Publisher v4.0.9 interface. The main window displays a table of variables with the following columns: Número, Nombre, Etiqueta, Ancho, InicioCol, FinalCol, Registro, Decimales, and Tipo de Dato. The table lists variables v40 through v47, with v44 highlighted. Below the table, the 'Documentación' tab is active, showing a tree view with 'Imputación' selected. A blue box highlights the 'Imputación' entry in the tree view. Below the tree view, there is a text area for the description of the element, which contains the text: 'Descripción del elemento: Describir el proceso para imputar o reemplazar datos inconsistentes o perdidos.'

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v40	P42257	g. Limpieza y mantenimiento del hogar.	1	62	62	1	0	Numérico(a)
v41	P42258	h. Otros	1	63	63	1	0	Numérico(a)
v42	P423	¿Por qué razón principal <...> realiza o colabora en	1	64	64	1	0	Numérico(a)
v43	P405	¿Por qué razón principal trabaja <...>?	1	65	65	1	0	Numérico(a)
v44	P408	En este trabajo <...> es:	1	66	66	1	0	Numérico(a)
v45	P409	Antes de descuentos cuánto ganó ...el mes pasado	7	67	73	1	0	Numérico(a)
v46	P410	Además del salario en dinero, ¿el mes pasado recibió	1	74	74	1	0	Numérico(a)
v47	P41051	¿En cuánto estima lo que recibió?	6	75	80	1	0	Numérico(a)

Fuente: DANE.

Recodificación y derivación: describe el proceso para derivar variables, aplica a los datos que se obtuvieron por recodificación de variables recolectadas, o mediante el cálculo de nuevas variables que no fueron obtenidas directamente en la recolección de datos.

Gráfica 65. Recodificación y derivación

The screenshot shows the Nesstar Publisher v4.0.9 interface. On the left is a project tree for 'MTI 2016'. The main window displays a 'Variables' table with columns: Número, Nombre, Etiqueta, Ancho, InicioCol, FinalCol, Registro, Decimales, and Tipo de Dato. The table lists variables v40 through v47. Variable v44 is highlighted in black. Below the table, the 'Documentación' tab is active, showing a tree structure with 'Recodificación y derivación' selected. A blue box highlights this element. Below the tree, a description box contains the following text:

Descripción del elemento:
Describir el proceso para derivar variables. Este elemento aplica a los datos que se obtuvieron por recodificación variables recolectadas, o mediante el cálculo de nuevas variables que no fueron obtenidas directamente en la recolección de datos.

Fuente: DANE.

Otros: incluye información adicional relacionada con la variable, como seguridad y notas.

Seguridad: describe el nivel del acceso apropiado para una variable.

Notas: puede ser utilizado para indicar alguna otra información acerca de la variable que no ha sido documentado en otro campo.

Gráfica 66. Otros

The screenshot shows the 'Variables' table in the Nesstar Publisher software. The table lists variables with their IDs, names, labels, widths, and data types. Below the table, there are sections for 'Documentación' (Documentation) and 'Seguridad' (Security).

Número	Nombre	Etiqueta	Ancho	InicioCol	FinalCol	Registro	Decimales	Tipo de Dato
v40	P42257	g. Limpieza y mantenimiento del hogar.	1	62	62	1	0	N Numérico(a)
v41	P42258	h. Otros	1	63	63	1	0	N Numérico(a)
v42	P423	¿Por qué razón principal <...> realiza o colabora en	1	64	64	1	0	N Numérico(a)
v43	P405	¿Por qué razón principal trabaja <...>?	1	65	65	1	0	N Numérico(a)
v44	P408	En este trabajo <...> es:	1	66	66	1	0	N Numérico(a)
v45	P409	Antes de descuentos ¿cuánto ganó...el mes pasado	7	67	73	1	0	N Numérico(a)
v46	P410	Además del salario en dinero, ¿el mes pasado recibí	1	74	74	1	0	N Numérico(a)
v47	P41051	¿En cuánto estima lo que recibió?	6	75	80	1	0	N Numérico(a)

Documentación

Estadísticas Ponderadores Documentación

Seguridad

El acceso a microdatos y mam-up se considera como de tratamiento especial respecto a la reserva estadística por tanto estará sujeto a la legislación que para el efecto defina el comité de aseguramiento de la reserva estadística. Resolución 173 de 2008.

Jerarquía de la Categoría

- 1 - Obrero o emplead
- 2 - Obrero o emplead
- 3 - Empleado domésti
- 4 - Trabajador por c
- 5 - Patrón o emplea
- 6 - Trabajador fami
- 7 - Trabajador fami
- 8 - Jornalero o pedr
- 9 - Otro

Información de la variable

Tipo de Datos: Numérico(a)

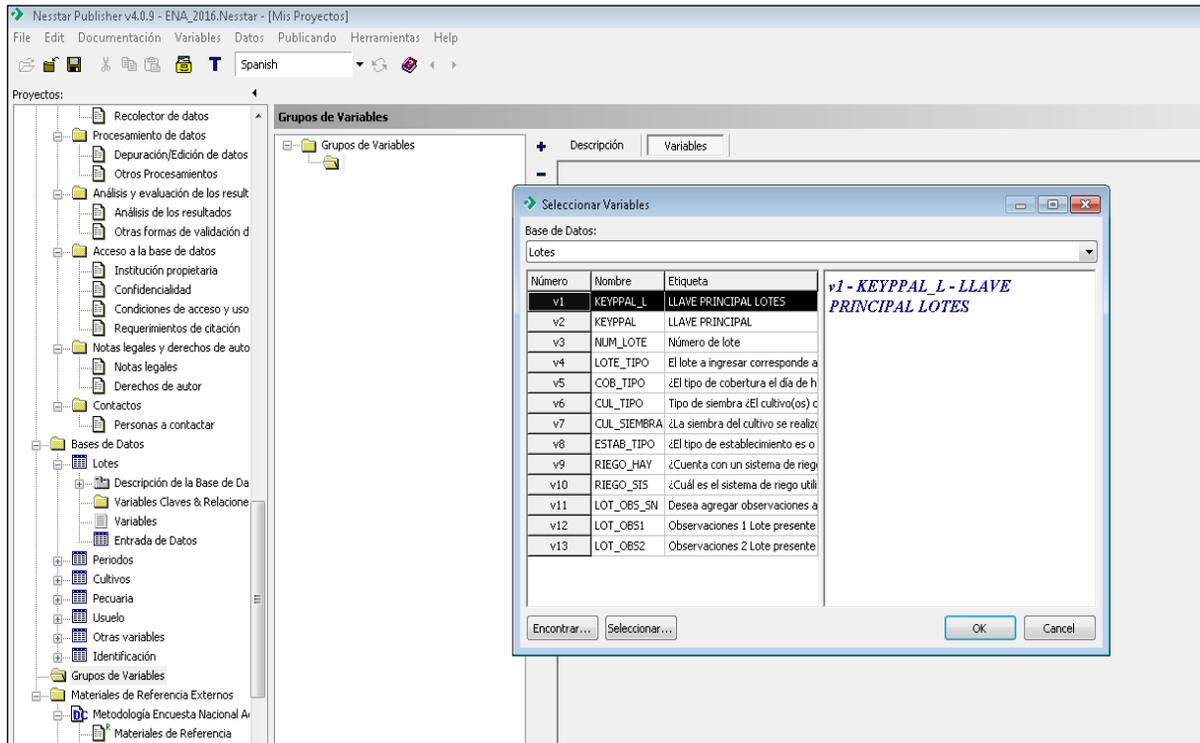
Medida: Nominal

Fuente: DANE.

2.4 Grupo de variables

Los archivos de datos pueden incluir cientos de variables. El software 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 los artículos que están conceptualmente conectados en una base de datos. Una variable puede pertenecer a más de un grupo y un grupo de variables puede incluir variables de más de un archivo de datos.

Gráfica 67. Grupo de variables



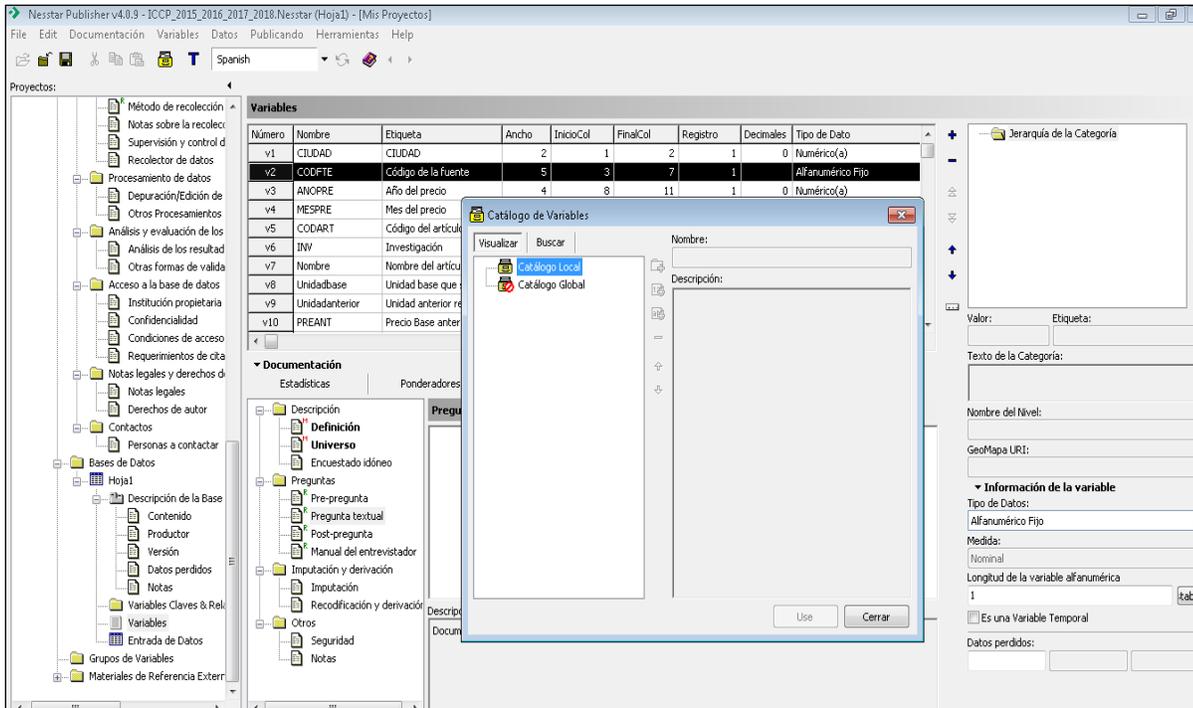
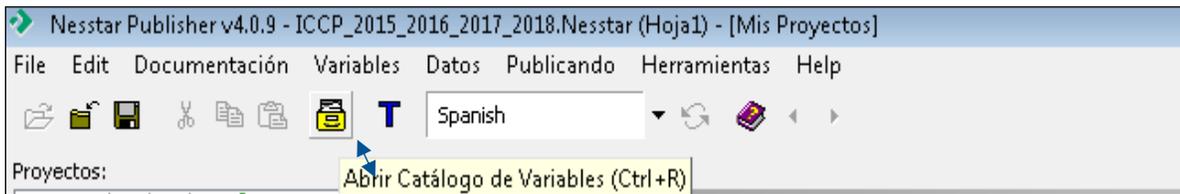
Fuente: DANE.

Para organizar las variables en grupo se debe crear el grupo y luego agregar las variables al grupo.

Solo dos archivos se pueden crear:

1. El catálogo local: es guardado en el computador del usuario y no puede ser ajustado por alguien más.
2. El catálogo global: se encuentra ubicado en un directorio compartido (como en una red local) facilitando la colaboración de las personas que se encuentran documentando estos datos.

Gráfica 68. Catálogo de variables



Fuente: DANE.

Documentación de la estructura de la base de datos (Diccionario de datos)

La información para realizar el proceso de documentación que se obtiene del diccionario de datos es: el código de la variable, la etiqueta o la descripción de la variable, el ancho o la longitud y la característica de la variable.

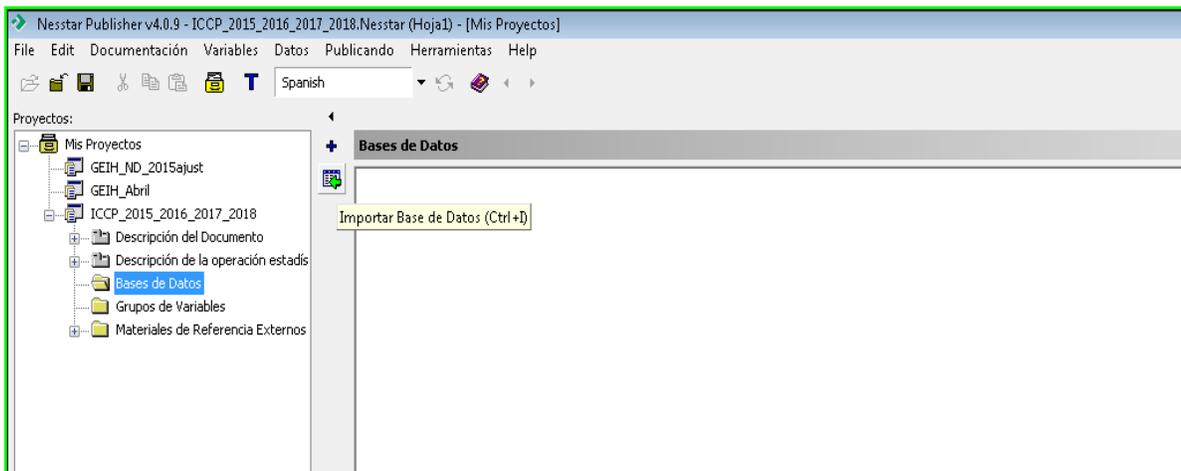
Teniendo en cuenta la estructura que se tiene en la hoja de cálculo, se realizan los siguientes pasos:

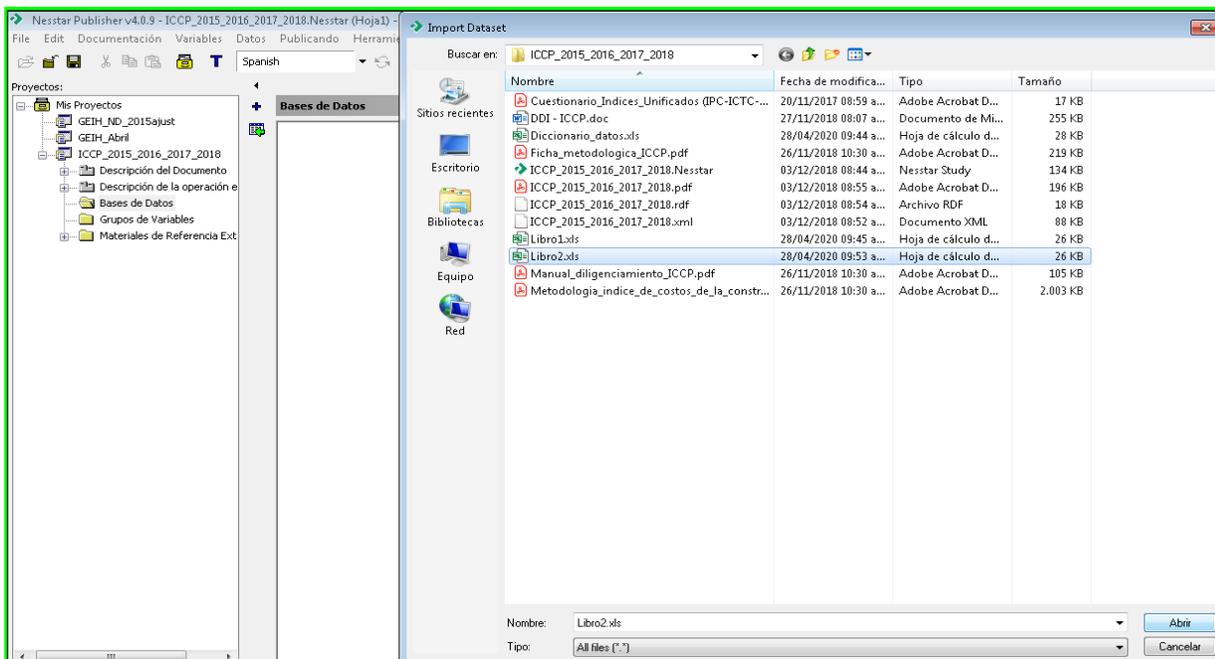
1. Seleccionar la columna de los códigos de las variables, copiar y en una hoja de cálculo nueva haciendo click derecho y seleccionar pegado especial, luego transponer (T). Esto hace que estos datos carguen al sistema como registros "Como ustedes saben Excel es un formato plano y no es un formato de base de datos por eso realizamos el procedimiento de transponer la Información", por último, le doy guardar en Excel 97-2003 y cierro este archivo.

En la hoja de cálculo que se guardaron los registros de los códigos de las variables, verificar si la mayoría son numéricas, si es así, ubicarse en la segunda fila de la primera columna y colocar un cero únicamente en esa celda, esto sirve para que todos los registros se encuentren señalizados y carguen esas variables con la característica numérica, de lo contrario, si la mayoría de las variables que tengo en mi base de datos son alfanuméricas dejamos nuestros datos tal cual.

2. Desde el Nesstar se importa la base datos y en la parte inferior seleccionar formato Excel y hacer click en abrir.

Gráfica 69. Importar bases de datos

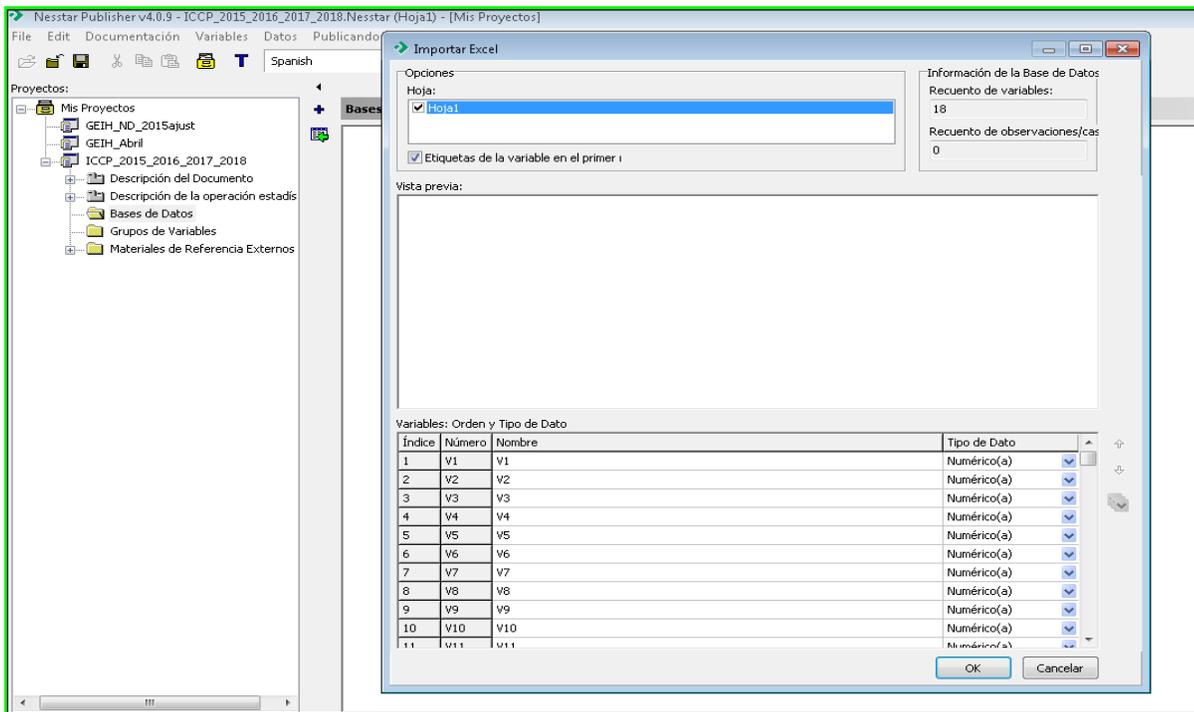




Fuente: DANE.

El aplicativo asume la primera hoja de cálculo donde tengo los datos y me indica el número de variables seleccionados; luego se selecciona el recuadro donde dice etiquetas de las variables. En la parte de vista previa observamos que todos los registros se encuentran en columnas y aparece el "cero (0)" en la segunda fila de la primera columna, teniendo en cuenta que la mayoría de las variables que voy a cargar en este ejemplo son numéricas, en la parte de abajo las observo con esta característica que son numéricas para terminar dar clic en Ok.

Gráfica 70. Nombre o código de la variable

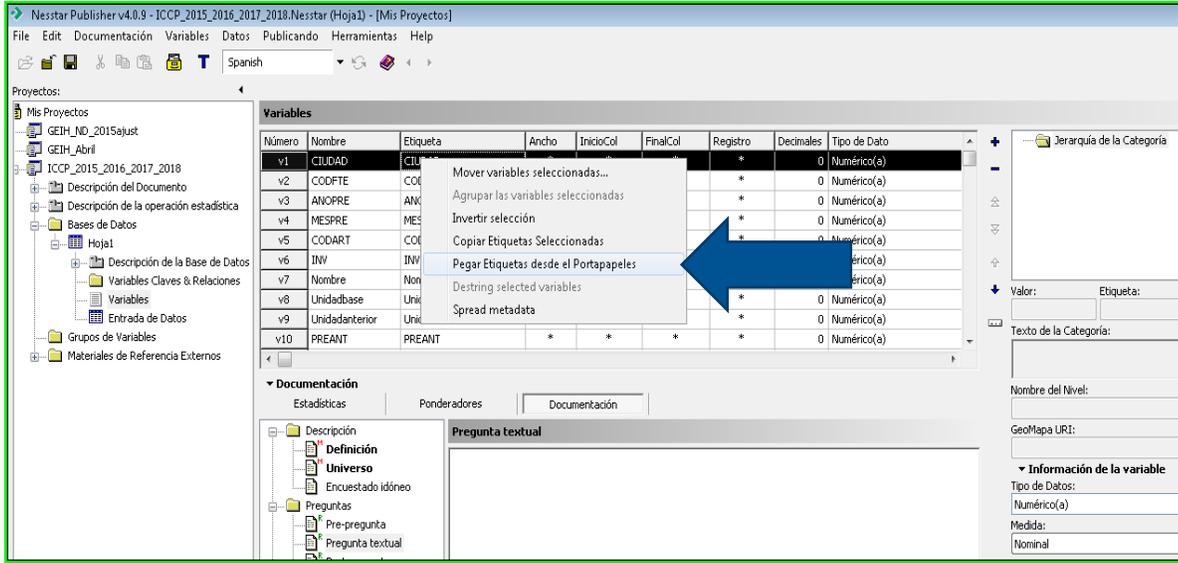


Fuente: DANE.

Para ver la estructura de base de datos selecciono el campo variable, observo que los códigos de las variables que cargue se encuentran en la primera columna llamada nombre.

Del diccionario de datos, copio la descripción de las variables, y en el Nesstar me ubico en la fila de la primera variable (queda sombreado), sobre está doy clic derecho y selecciono pegar etiquetas desde el portapapeles.

Gráfica 71. Etiquetas



Fuente: DANE.

El proceso de documentación del ancho de las variables se realiza una a una con los valores que tengo en el diccionario de datos.

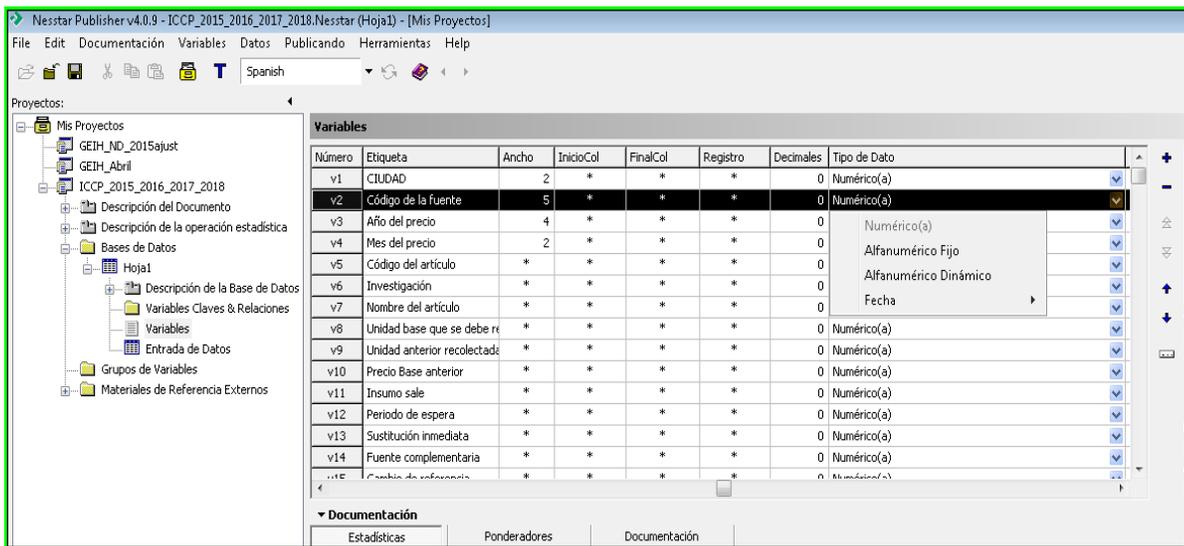
Gráfica 72. Documentación ancho de las variables



Fuente: DANE.

En la sección tipo de dato, seleccionamos la característica de cada una de las variables (numéricas, alfanuméricas y/o fecha), éstas las puedo cambiar con la pestaña desplegable que se encuentra al lado derecho.

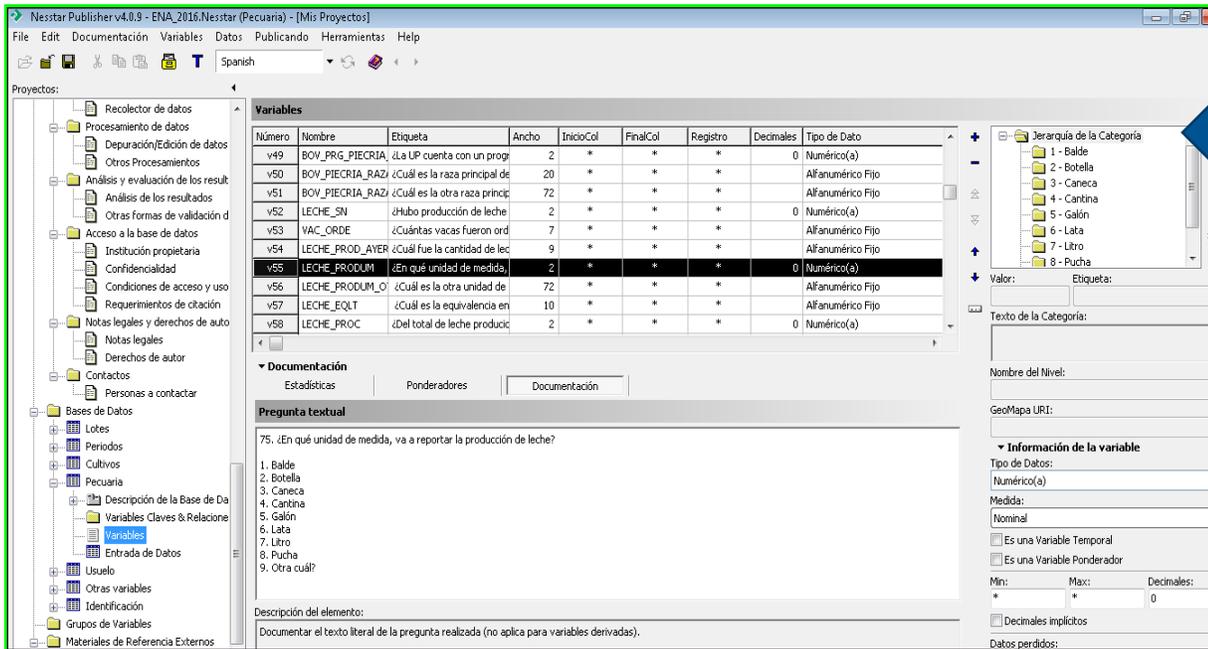
Gráfica 73. Tipo de dato



Fuente: DANE.

Para esta parte del proceso de documentación, se debe verificar por medio del formulario si las variables tienen categorías. Si es así, se realiza el proceso de documentación de la pregunta con sus respectivas categorías en el campo de pregunta textual. Y luego se crean las categorías en la pestaña Jerarquía de la categoría.

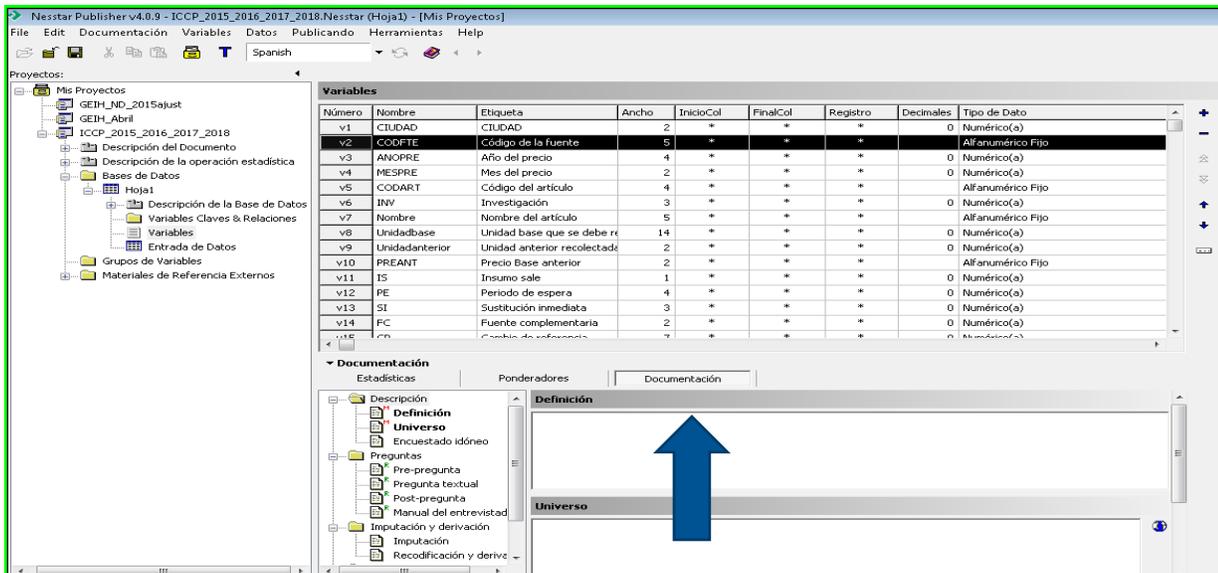
Gráfica 74. Categorías



Fuente: DANE.

El proceso de documentación de cada una de las variables se realiza como en el caso uno Base de datos con microdatos en elementos de la documentación.

Gráfica 75. Elementos de la documentación



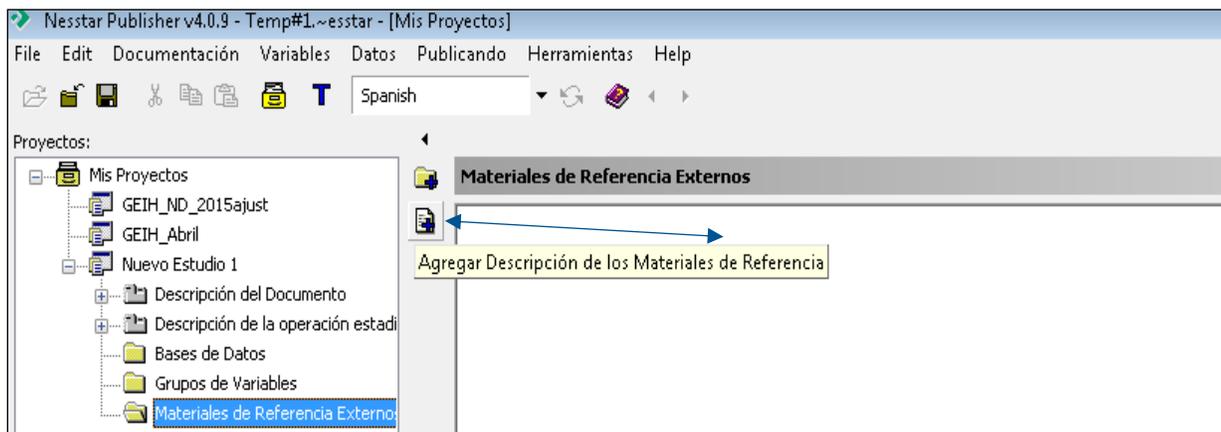
Fuente: DANE.

2.5 Materiales de referencia externos

Esta sección se utiliza para proporcionar un listado y una descripción de materiales como: documentos (manuales, cuestionarios, metodología, ficha metodológica). En esta se emplea el estándar estadístico Dublin Core.

El primer paso para documentar el **Materiales de Referencia Externos** es agregar una descripción de los materiales de referencia para la metodología, la ficha metodológica y el cuestionario de la operación estadística.

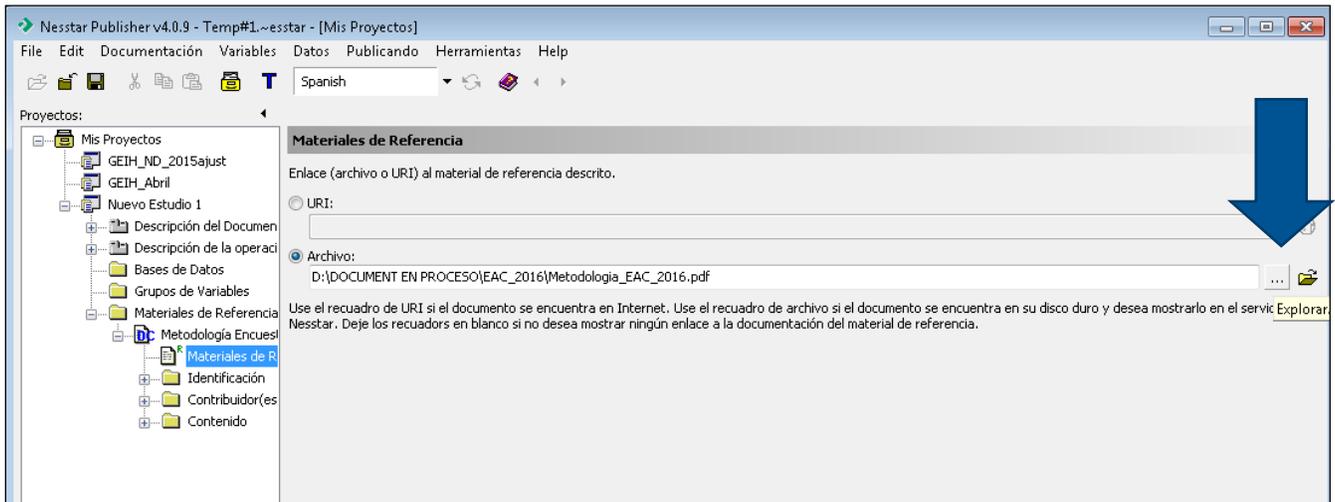
Gráfica 76. Agregar descripción de los materiales de referencia



Fuente: DANE.

Se escribe el nombre del documento y se carga en **Explorar**.

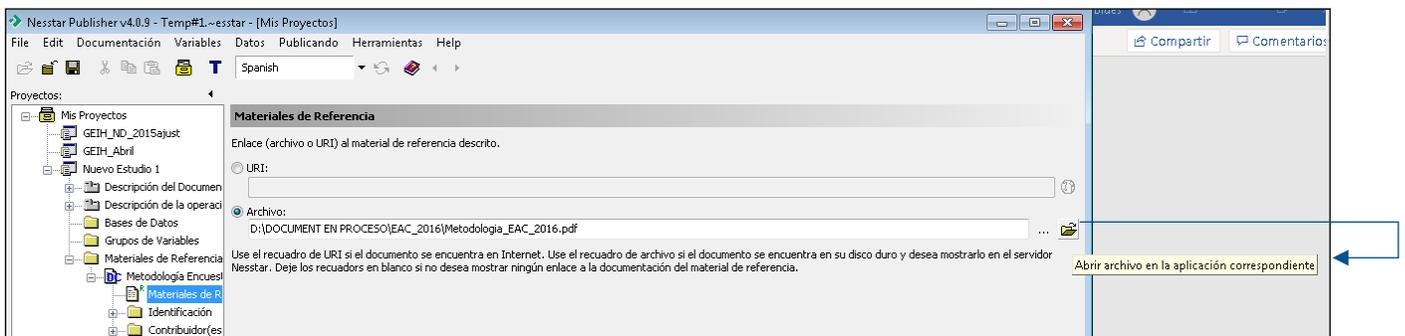
Gráfica 77. Cargue del documento



Fuente: DANE.

Se debe verificar el documento enlazado en el material de referencia dando clic en **abrir archivo en la aplicación correspondiente**.

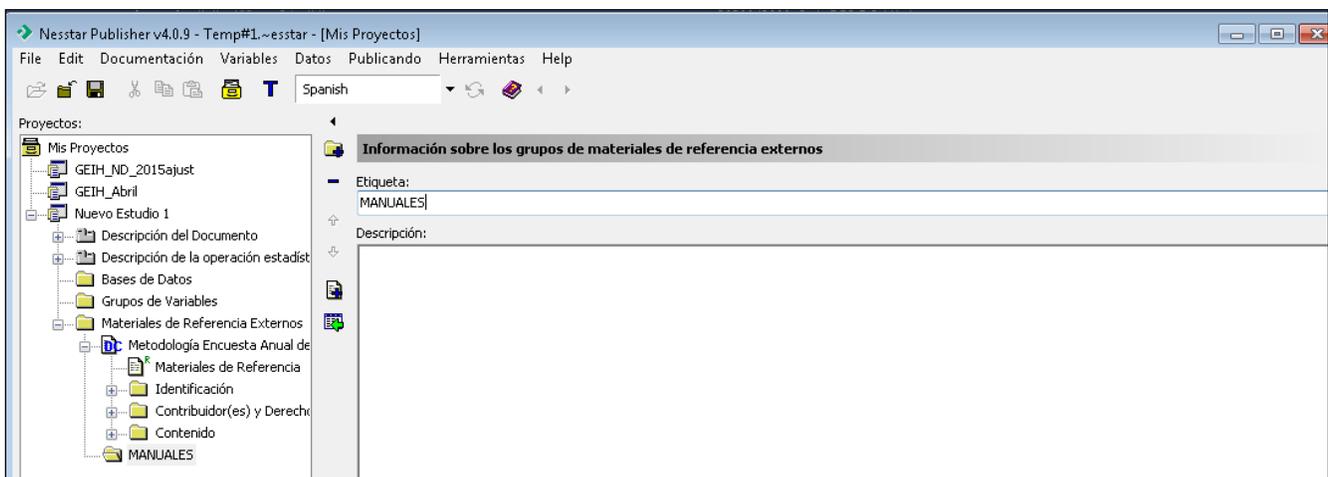
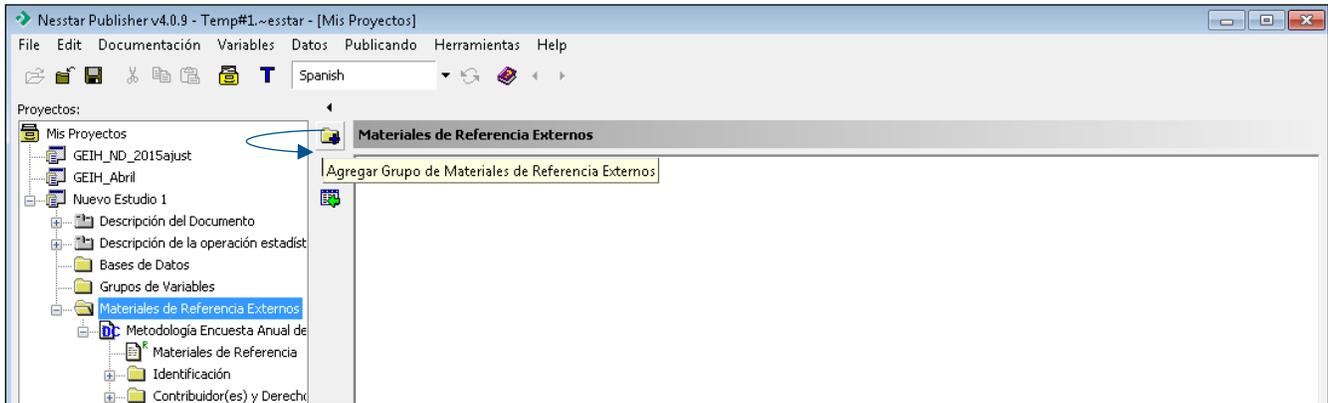
Gráfica 78. Abrir archivo en la aplicación correspondiente



Fuente: DANE.

Agregar grupo de materiales de referencia externos: corresponde a la creación de carpetas para incluir manuales, metodologías, guías, etc.

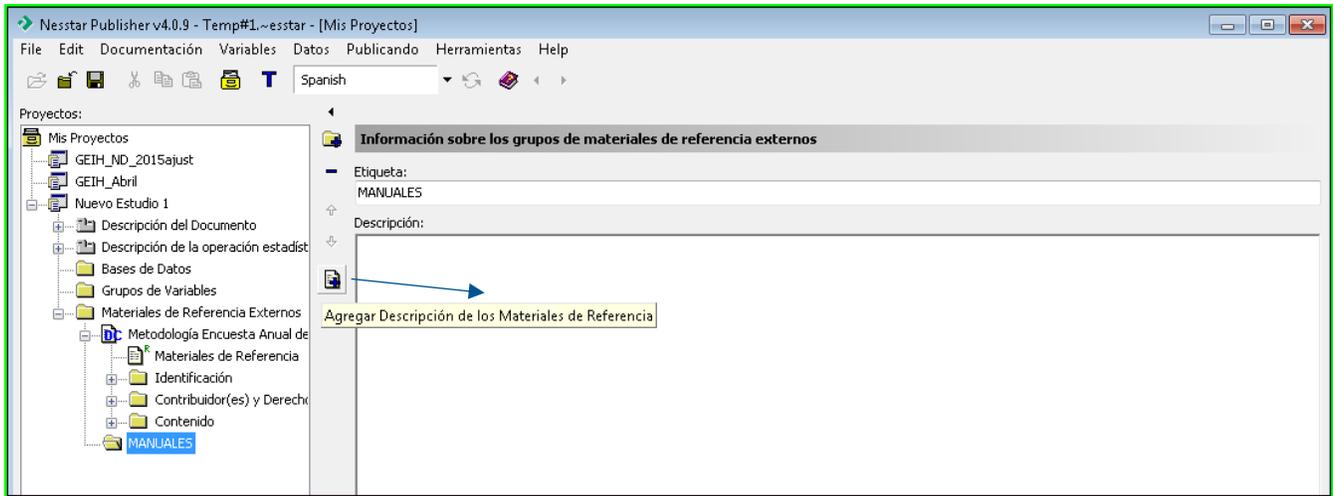
Gráfica 79. Agregar grupo de materiales de referencia externos



Fuente: DANE.

Luego hay que ubicarse sobre el título del documento del grupo y agrego la descripción de los materiales de referencia.

Gráfica 80. Agregar descripción de los materiales de referencia

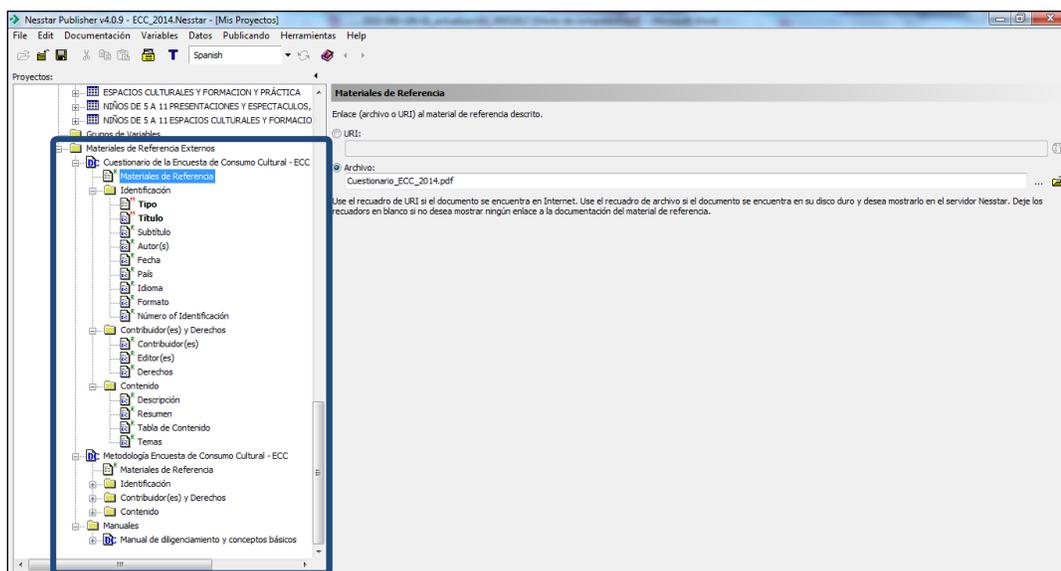


Fuente: DANE.

El segundo paso para documentar **Material de Referencia Externo** se documenta el contenido para manuales, metodologías, guías, etc.

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:

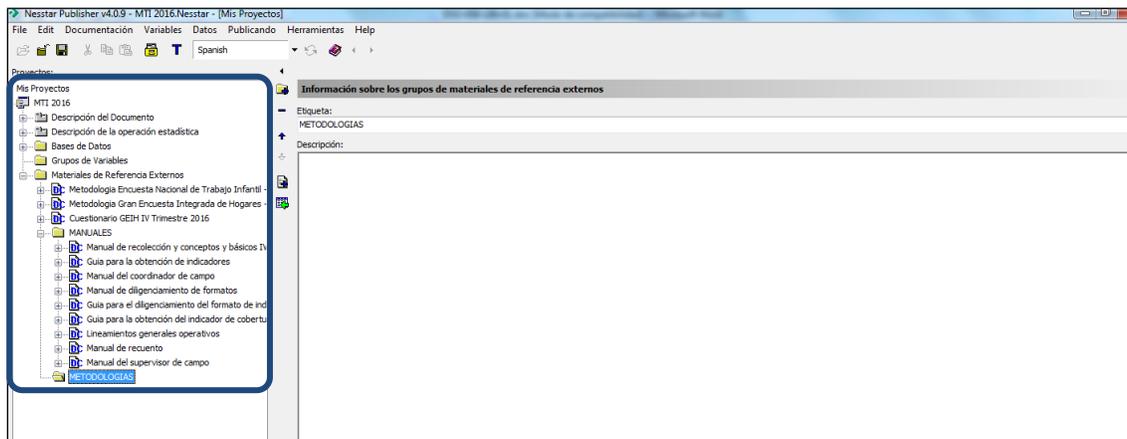
Gráfica 81. Ejemplo de presentación de material de referencia externo



Fuente: DANE.

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 82. Ejemplo de presentación de los nombres de los documentos

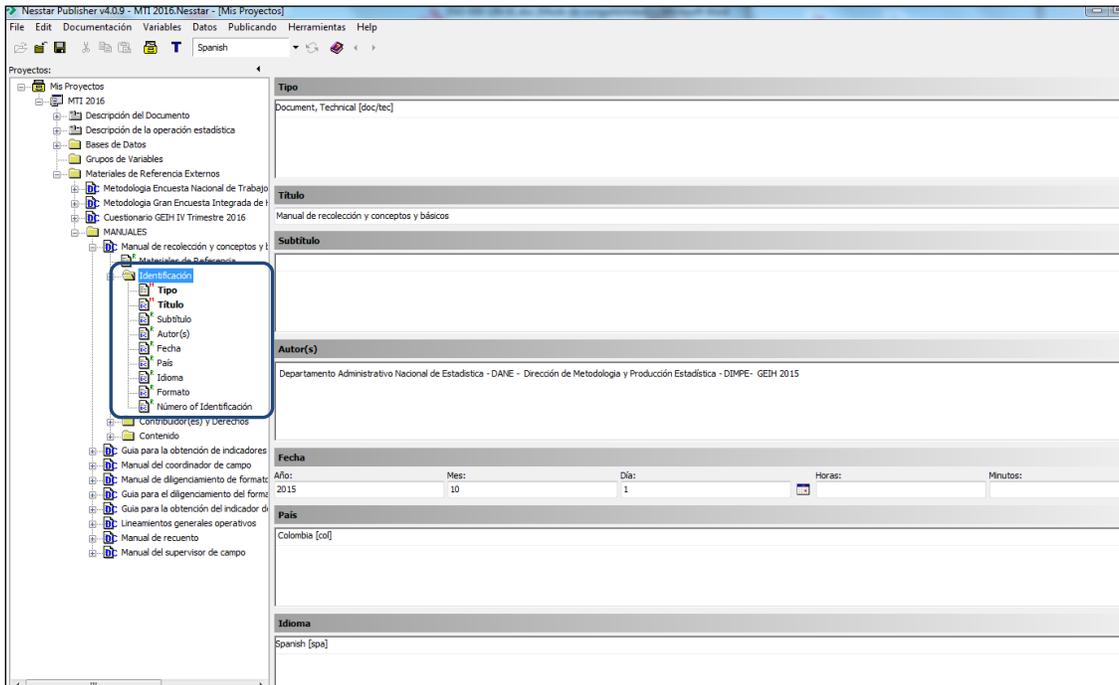


Fuente: DANE.

IDENTIFICACIÓN

En este grupo se relacionan los elementos con los que se identifica el material relacionado, con el objeto de dar claridad frente al tipo de documentos que se encontrarán. Consta de los siguientes elementos: tipo, título, subtítulo, autor, fecha, país, idioma, formato y número de identificación.

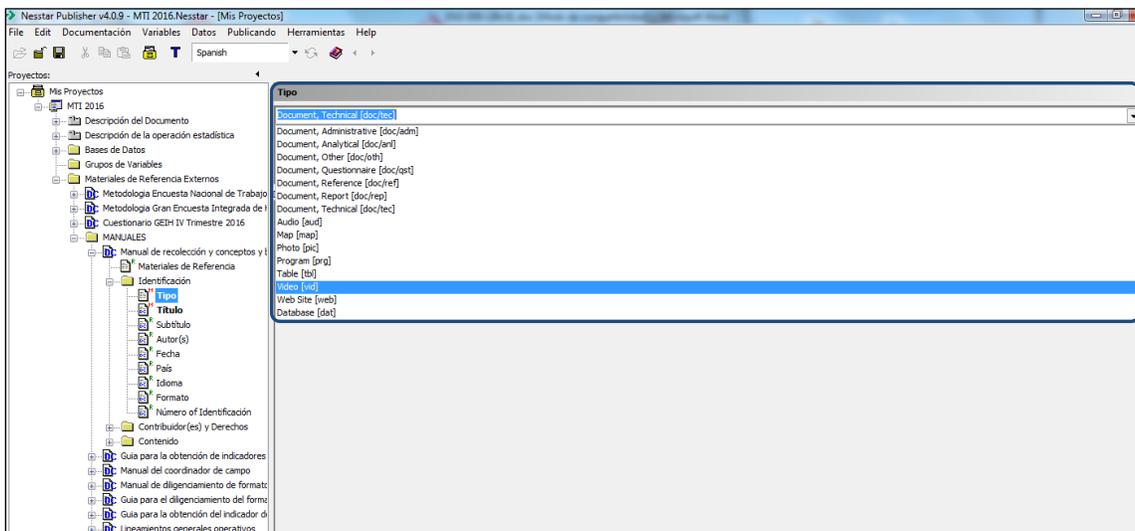
Gráfica 83. Identificación



Fuente: DANE.

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

Gráfica 84. Tipo de material de referencia



Fuente: DANE.

Título y subtítulo: en estos elementos se debe ingresar un nombre formal y otro de manera opcional que es secundario al material de referencia.

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.

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.

País: en este elemento se listan todos los países dentro del alcance del material de referencia en el elemento.

Idioma: en este elemento se listan todos los idiomas en que aparecen los materiales de referencia.

Formato: se debe seleccionar un objeto del listado para identificar el formato del archivo del material de referencia.

Número de identificación: si existe se debe presentar el identificador del material de referencia.

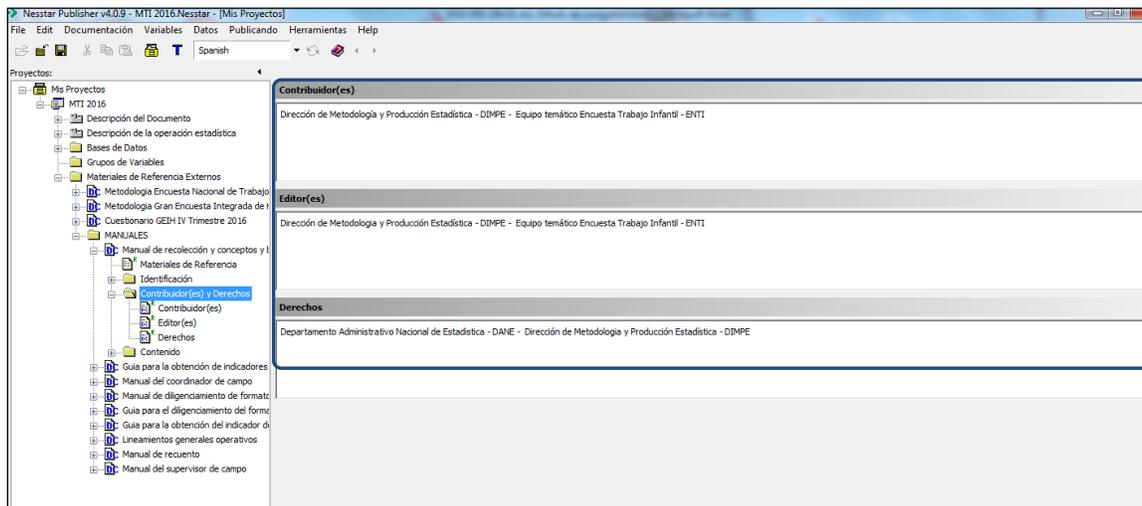
CONTRIBUIDOR(ES) Y DERECHOS

En este elemento es necesario citar a las personas y las organizaciones que han apoyado o contribuido al desarrollo del material de referencia.

Editor: organización o persona responsable de la edición del documento y/o difusión.

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

Gráfica 85. Contribuidor(es) y derechos legales

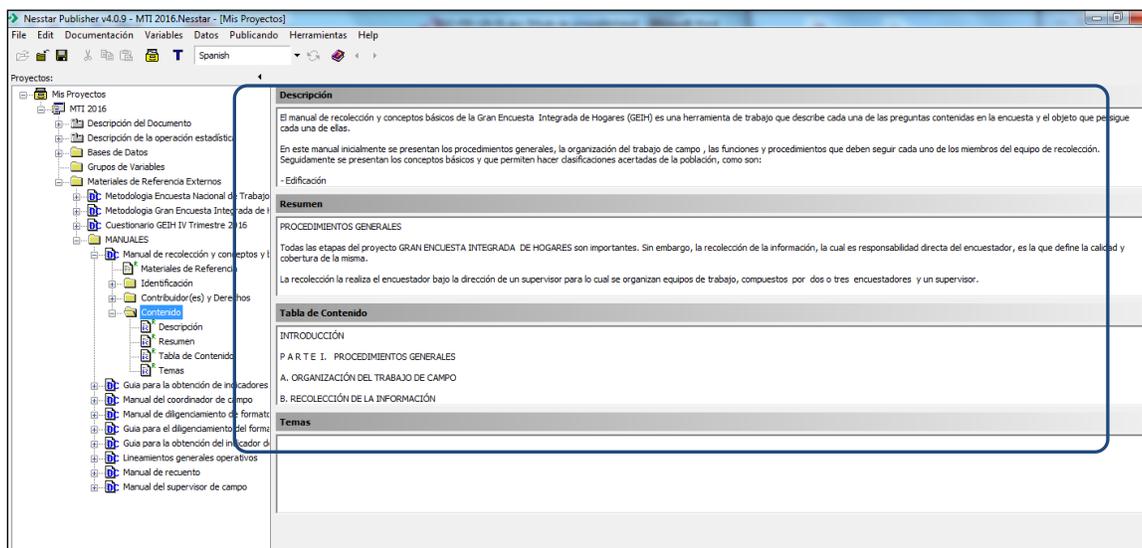


Fuente: DANE.

CONTENIDO

En este elemento se debe describir brevemente el contenido del material de referencia: descripción, resumen, tabla de contenido y temas.

Gráfica 86. Contenido



Fuente: DANE.

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, el cuestionario y otro documento.

Temas: este elemento es usado para describir los temas abordados en el material de referencia.

3. Archivo Nacional de Datos (ANDA)

Luego de realizar el proceso de documentación a través de los estándares estadísticos descritos y el software Nesstar Publisher, los microdatos y los metadatos de las operaciones estadísticas del SEN reposarán en el ANDA⁸, que es un sistema de consulta, basado en catálogos, desarrollado por la Red Internacional de Encuestas a Hogares (IHSN por sus siglas en inglés) con el respaldo del Banco Mundial y la OCDE.

El objetivo del ANDA es ayudar a los productores en la preservación de datos y en la difusión de estos, para consultas con fines de investigación u otros usos.

The screenshot shows the ANDA website interface. At the top, there is a header with the ANDA logo and the text 'Acceso a Metadatos y Microdatos anonimizados'. Below the header, a paragraph states: 'Con el ánimo de mejorar el acceso a Datos Abiertos, hemos actualizado el diseño de la interfaz del Archivo Nacional de Datos (ANDA). En ésta, se agrupan las Operaciones Estadísticas en tres categorías: **Economía, Sociedad y Territorio.**'

Below this, there is a section titled 'Consideraciones para el uso de la información' with a note: 'Las bases anonimizadas y disponibles en el ANDA han surtido un proceso de modificación y transformación de los datos originales, con el objetivo de garantizar la confidencialidad de las unidades de análisis, por lo cual es necesario tener en cuenta estas modificaciones para el uso de la información.'

The main content area features a card for 'Estadísticas de exportaciones'. To the left of the card, it says: 'Ésta es la forma en como se presenta la información de cada operación, de acuerdo con cada categoría:'. The card itself has two tabs: 'Metadatos' (with the date 'Marzo 5 de 2019') and 'Microdatos' (with the date 'Marzo 5 de 2019'). To the right of the card, there are three labels: 'Nombre de la operación estadística', 'Última actualización de Microdatos', and 'Última actualización de Metadatos'.

At the bottom of the card, it says: 'Consulta las operaciones en cada una de las áreas temáticas'. Below this, there are three buttons: 'Economía', 'Sociedad', and 'Territorio'.

⁸ Para mayor información: http://formularios.dane.gov.co/Anda_4_1/index.php/home

Bibliografía

- DANE (20012a). Guía para documentar y codificar documentación técnica. Sistema Integrado de Gestión Institucional. Bogotá: DANE.
- DANE (2012b). Guía para documentar y codificar documentación técnica. Sistema Integrado de Gestión Institucional. Bogotá: DANE.
- DANE (2010). Instructivo para la elaboración de guías. Sistema Integrado de Gestión Institucional. Bogotá: DANE.
- DANE (s.f.). Instructivo para la elaboración de guías. Sistema Integrado de Gestión Institucional. Bogotá: DANE.
- DANE (2012c). 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>