

Departamento Administrativo Nacional de Estadística



Diseño DSO

Dirección de Metodología y Producción Estadística / DIMPE

MANUAL ANONIMIZACIÓN DE LAS BASES DE QUINTO CENSO NACIONAL DE ARROZ MECANIZADO 2023

Agosto 2024

	MANUAL ANONIMIZACIÓN DE LAS BASES DE QUINTO CENSO NACIONAL DE ARROZ MECANIZADO 2023	CÓDIGO: DSO- ENA-MOT-001 VERSIÓN: 2 FECHA: 23/Ago/20 24
PROCESO: Diseño	OPERACIÓN ESTADÍSTICA: QUINTO CENSO NACIONAL DE ARROZ MECANIZADO 2023	

TABLA DE CONTENIDO

- 1. INTRODUCCIÓN**
- 2. OBJETIVO**
- 3. DEFINICIONES**
- 4. CONTENIDO**
 - 4.1. PROTOCOLO DE ANONIMIZACIÓN DE LAS BASE – QUINTO CENSO NACIONAL DE ARROZ MECANIZADO**
 - 4. 1. 1. VARIABLES QUE DEBEN EXCLUIRSE DE LA BASE**
 - 4. 1. 2. VARIABLES QUE DEBEN SER RECODIFICADAS**
 - 4. 1. 3. VARIABLES CONTINUÚAS CATEGORIZADAS**
 - 4. 1. 4. DISPOSICIÓN DE DATOS**
 - 4. 1. 5. APLICACIÓN DE LINEAMIENTOS PARA ANONIMIZACIÓN**
- 5. REGISTROS**
- 6. BIBLIOGRAFIA**
- 7. ANEXOS**

1. INTRODUCCIÓN

Mediante Resolución 1503 de 2011, el DANE establece medidas de orden técnico, administrativo y jurídico para garantizar la seguridad de los datos, archivos y/o bases de datos que contienen información de personas naturales y jurídicas.

El Departamento Administrativo Nacional de Estadística DANE ofrece los niveles apropiados de Seguridad de la Información para la generación y el acceso a los Microdatos Anonimizados. Las metodologías de anonimización son definidas por los diferentes equipos temáticos y la determinación de acceso tomada por el comité de aseguramiento de la Reserva Estadística de acuerdo con el protocolo definido.

La Oficina de Sistemas está encargada de aplicar dicha metodología para generar estos archivos anonimizados, los cuales son almacenados en servidores ubicados en el centro de cómputo en un ambiente seguro, haciendo uso de la infraestructura de seguridad implementada y aplica los niveles de acceso definidos a través de autenticación, autorización y registro de las actividades, ofreciendo la integridad, disponibilidad y confidencialidad apropiadas.

A continuación, se presenta la metodología de anonimización para los Microdatos del Quinto Censo Nacional Arrocero 2023.

2. OBJETIVO

Presentar la metodología para realizar el proceso de anonimización de los Microdatos del Quinto Censo Nacional Arrocero año 2023, estableciendo las variables que deben retirarse de las bases de datos y las que deben ser recodificadas con el fin de cumplir con la normatividad de seguridad de la información y reserva estadística.

3. DEFINICIONES

Área cosechada: Área total en la que se ha recolectado el cultivo; se excluye el área destruida por sequías, inundaciones, plagas o cualquier otra razón. (DANE – Conceptos estandarizados)

Arroz mecanizado: Cultivo de arroz en el cual se emplean máquinas (tractores, combinadas, aviones y sistemas de riego) para realizar una o varias labores del proceso productivo del cultivo; entre otros, preparación del suelo, siembra, control de malezas y plagas, fertilización o recolección. Este a su vez, se divide en dos sistemas de producción, arroz riego y arroz seco. (DANE – Conceptos estandarizados)

Arroz riego: producción del arroz en el que el productor suministra agua al cultivo mediante sistemas de riego. (DANE – Conceptos estandarizados)

Arroz seco: Forma de producción del arroz en la que el agua que requiere el cultivo proviene de la lluvia. (DANE – Conceptos estandarizados)

Área sembrada: Área cultivada con una determinada planta o asocio de plantas en forma compacta para ser cosechada durante los períodos de referencia de la encuesta. (DANE – Conceptos estandarizados).

Finca arroceras: Superficie de tierra continua geográficamente dedicada total o parcialmente al cultivo de arroz. (DANE – Conceptos estandarizados)

Población Objetivo. La población objetivo está constituida por la totalidad de los productores arroceros y las Unidades Productoras de Arroz donde se desarrolle la actividad de siembra o cosecha de arroz mecanizado en la zona rural del territorio nacional durante 2023.

Uso del suelo en la finca: Uso actual y efectivo de la superficie de la finca. Incluye la superficie total de las tierras que se emplean o no para la producción agropecuaria. (DANE – Conceptos estandarizados)

Universo de estudio: El universo del 5° CNAM está constituido por el área rural del país donde se desarrolle la actividad de siembra o cosecha de arroz mecanizado durante 2023.

Zona arroceras: Región geográfica que agrupa las zonas con condiciones económicas y agronómicas similares donde se cultiva arroz. (DANE – Conceptos estandarizados)

4. CONTENIDO

4.1. PROTOCOLO DE ANONIMIZACIÓN DE LA BASE – 5° CNAM

El Censo Nacional Arroceros, tiene como objetivo, "proporcionar información estratégica y actualizada sobre los productores y las Unidades Productoras de Arroz donde se desarrolle la actividad de siembra y/o cosecha de arroz riego y secano mecanizado en Colombia durante el año 2023, con el fin de apoyar la toma de decisiones en los ámbitos económico, social y ambiental del sector arroceros. Este se realiza mediante entrevista directa a los productores arroceros."

El 5° CNAM, a diferencia de los cuatro censos anteriores realizados en el país, consideró además de la información sobre área cosechada, producción y rendimiento de arroz mecanizado, variables que permiten caracterizar aspectos socioeconómicos de los agricultores de arroz mecanizado, así como, conocer el manejo de los residuos de los productos químicos de uso agrícola que se utilizan.

El levantamiento se desarrolló mediante dos operativos de recolección de datos en campo, correspondientes a la actividad del cultivo de arroz mecanizado que se realizó en el país durante el primer y segundo semestre de 2023.

La identificación de las necesidades de información para el Quinto Censo Nacional de Arroz Mecanizado 2023 (5° CNAM) inició con la recopilación de solicitudes de necesidades de información estadística tanto al gremio arroceros (Fedearroz-FNA) como al DANE por parte de los diferentes usuarios de la información.

A continuación, se presenta la metodología de anonimización para el 5° CNAM, en el cual, se especifican los diferentes métodos aplicados según los tipos de variables y el riesgo que tienen para poder identificar las fuentes de información. Entre estos métodos se encuentran los de exclusión, recodificación y recategorización variables.

4.1.1. VARIABLES QUE DEBEN EXCLUIRSE DE LA BASE

Con el fin de garantizar la reserva estadística de las fuentes y que no sea posible deducir de alguna información de carácter individual, se deben eliminar de la base que se dispondrá al público varios tipos de variables, entre ellas:

- Las variables de identificación de: la finca, la UPA, los productores o de quien suministre la información, cuando no lo hace el productor directamente, (administradores, familiares, gerentes, etc.). Esto incluye datos de coordenadas geográficas, códigos catastrales, nombres, números de identificación, direcciones, números de teléfono, etc.
- De igual forma se eliminan variables operativas como fechas del día de la entrevista, del cargue y envío de información, nombres y códigos del personal de recolección y de los encuestados, etc.
- Otras variables que se eliminan son: el nombre y código del municipio. De esta manera se evita que sea posible la identificación de un productor a partir del cruce de varias variables. Se realizaron varias pruebas y ejercicios para evitar eliminar la variable de municipio y que los usuarios puedan consultar esta información pero dado que en muchos municipios la cantidad de productores es menor a cuatro(4) productores, no fue posible lograr una estrategia de anonimización que no viole la reserva estadística, por ello fue necesario eliminarla de la base.

En las tablas 1, 2 y 3 se listan las variables que se excluyen de la base de datos según la tabla a la que pertenecen en la base de datos.

Tabla 1. Variables a excluirse tabla finca

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_1_FINCAS	NO	NOMENCU	Código del encuestador(a)
CNAV_1_FINCAS	NO	COORDINA	2.1. Nombre del coordinador
CNAV_1_FINCAS	NO	COORDENADAS_MANUALES	2. Punto de ubicación
CNAV_1_FINCAS	NO	COORDENADAS_LATITUDE	2.1. Coordenada geográfica de la finca => Latitud
CNAV_1_FINCAS	NO	COORDENADAS_LONGITUDE	2.1. Coordenada geográfica de la finca => Longitud
CNAV_1_FINCAS	NO	COORDENADAS_ALTITUDE	2.1. Coordenada geográfica de la finca => Altitud
CNAV_1_FINCAS	NO	COORDENADAS_ACCURACY	2.1. Coordenada geográfica de la finca => Exactitud
CNAV_1_FINCAS	NO	ENC_NOM	¿Nombre(s) y apellido(s) de la persona que suministra la información?
CNAV_1_FINCAS	NO	ENC_TIPOTEL_OPA	8. Número telefónico de la persona que suministra la información => 1. Teléfono fijo
CNAV_1_FINCAS	NO	ENC_TIPOTEL_OPB	8. Número telefónico de la persona que suministra la información => 1. Teléfono móvil
CNAV_1_FINCAS	NO	ENC_TIPOTEL OPC	8. Número telefónico de la persona que suministra la información => 2. No tiene
CNAV_1_FINCAS	NO	ENC_TIPOTEL OPD	8. Número telefónico de la persona que suministra la información => 3. No informa
CNAV_1_FINCAS	NO	ENC_TELFIJO	Número telefónico fijo
CNAV_1_FINCAS	NO	ENC_TELMOVIL	Número móvil
CNAV_1_FINCAS	NO	MPIO	Nombre del municipio
CNAV_1_FINCAS	NO	VEREDA	Nombre de la vereda
CNAV_1_FINCAS	NO	VEREDA_OTC	Otra

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_1_FINCAS	NO	FINCA_NOM	Nombre la finca
CNAV_1_FINCAS	NO	FINCA_PROP	Nombre del(la) propietario(a) de la finca
CNAV_1_FINCAS	NO	DISTRI_PREDIO	¿Cuál es el Código del Predio dentro del Distrito de Riego?
CNAV_1_FINCAS	NO	FEC_CREACION	Fecha de Creación de la Encuesta
CNAV_1_FINCAS	NO	FEC_CARGUE	Fecha de Cargue
CNAV_1_FINCAS	NO	KEYP	Llave UUID
CNAV_1_FINCAS	NO	OBS_SN	81. ¿Desea agregar observaciones a la encuesta?
CNAV_1_FINCAS	NO	OBS	81.1. Observaciones a la encuesta
CNAV_1_FINCAS	NO	AUTORIZACION	83. Señor(a) Supervisor: Digite la clave para autorizar el cierre del formulario Clave:
CNAV_1_FINCAS	NO	P_ENC_TERM	84. ¿La encuesta de la finca fue realizada satisfactoriamente?
CNAV_1_FINCAS	NO	FECHA_VIS	Fecha de visita
CNAV_1_FINCAS	NO	MES_VIS	Mes de la visita

Fuente: DANE- 5° CNAM

Tabla 2. Variables a excluirse tabla **productor**

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_5_PRODUCTORES	NO	CON_PROD	Contador productor
CNAV_5_PRODUCTORES	NO	FINCA_NUM	Formulario corresponde a la finca número
CNAV_5_PRODUCTORES	NO	PRODSOC_NOM	¿Nombre(s) y apellido(s) del productor(a)?
CNAV_5_PRODUCTORES	NO	PRODSOC_TI	¿Cuál es el tipo de documento de identificación del productor(a)?
CNAV_5_PRODUCTORES	NO	PRODSOC_ID	¿Cuál es el número de documento de identificación del productor(a)?
CNAV_5_PRODUCTORES	NO	PRODSOC_TIPOTEL_O PA	52. Número telefónico del (la) productor(a) => 1. Teléfono fijo => 1. Teléfono fijo
CNAV_5_PRODUCTORES	NO	PRODSOC_TIPOTEL_O PB	52. Número telefónico del (la) productor(a) => 1. Teléfono móvil => 1. Teléfono móvil
CNAV_5_PRODUCTORES	NO	PRODSOC_TIPOTEL_O PC	52. Número telefónico del (la) productor(a) => 2. No tiene => 2. No tiene
CNAV_5_PRODUCTORES	NO	PRODSOC_TIPOTEL_O PD	52. Número telefónico del (la) productor(a) => 3. No informa => 3. No informa

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_5_PRODUCTORES	NO	PRODSOC_TELFIJO	Número telefónico fijo
CNAV_5_PRODUCTORES	NO	PRODSOC_TELMOVIL	Número móvil
CNAV_5_PRODUCTORES	NO	EMP_NIT	Número de NIT de la empresa:
CNAV_5_PRODUCTORES	NO	EMP_DV	Cuál es el dígito de verificación
CNAV_5_PRODUCTORES	NO	EMP_NOMBRE	Nombre comercial de la empresa
CNAV_5_PRODUCTORES	NO	EMP_TIPOTEL_OPA	63. ¿Número telefónico empresa? => 1. Teléfono fijo
CNAV_5_PRODUCTORES	NO	EMP_TIPOTEL_OPB	63. ¿Número telefónico empresa? => 1. Teléfono móvil
CNAV_5_PRODUCTORES	NO	EMP_TIPOTEL_OPC	63. ¿Número telefónico empresa? => 2. No tiene
CNAV_5_PRODUCTORES	NO	EMP_TIPOTEL_OPD	63. ¿Número telefónico empresa? => 3. No informa
CNAV_5_PRODUCTORES	NO	EMP_TELFIJO	Número telefónico fijo
CNAV_5_PRODUCTORES	NO	EMP_TELMOVIL	Número móvil
CNAV_5_PRODUCTORES	NO	EMP_RLEG	¿Cuál es el nombre del (la) representante legal?
CNAV_5_PRODUCTORES	NO	FEC_CARGUE	Fecha de cargue

Fuente: DANE- 5° CNAM

Tabla 3. Variables a excluirse tabla UPA

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_2_UPAS	NO	CON_UPA	Numero de UPA

Fuente: DANE- 5° CNAM

4. 1. 2. VARIABLES QUE DEBEN SER RECODIFICADAS

Dado que la base de datos del 5° CNAM está conformada por varias tablas dependiendo de la unidad a trabajar, finca, UPA, productor, maquinaria, siembra, cosecha, uso del suelo, tendrá asociadas diferentes variables. Se debe tener un cuidado especial con las variables "LLAVE" que son las que permiten relacionar las diferentes tablas.

En la base original la llave que permite encadenar las diferentes tablas se construye a partir de la identificación de la finca, la cual a su vez esta codificada según el código del departamento, el municipio y un consecutivo al interior de cada municipio.

La metodología de recodificación consiste en transformar la variable anidada que actúa como llave para las diferentes tablas, por otras que desempeñen la misma función pero que no permitan identificar la finca en mapas al cruzarlo con información de otras fuentes. Esta transformación se realiza utilizando algoritmos para generar códigos consecutivos alfanuméricos ordenados

aleatoriamente. La oficina de sistemas conserva una correlativa entre las variables recodificadas y las originales.

En la tabla 4 se presenta la lista de variables a recodificar

Tabla 4. Variables recodificación

Nombre de la tabla	Llave Primaria	Id. de la variable	Descripción de la variable
CNAV_1_FINCAS	SI	KEY_COFINCA	Llave de la finca
CNAV_1_FINCAS	NO	ID_COFINCA	ID_COFINCA
CNAV_1_FINCAS	NO	COFINCA	Código de la finca
CNAV_1_FINCAS	NO	COFINCA_NEW	Código de la FINCA nueva
CNAV_2_UPAS	Si	KEY_COFINCAUPA	Llave UPA
CNAV_2_UPAS	NO	KEY_COFINCA	Llave de la finca
CNAV_3_SIEMBRAS	SI	KEY_COFINCAUPA_SIE	Llave siembra
CNAV_3_SIEMBRAS	NO	KEY_COFINCAUPA	Llave de la UPA
CNAV_4_COSECHAS	SI	KEY_COFINCAUPA_CO S	Llave Cosecha
CNAV_4_COSECHAS	NO	KEY_COFINCAUPA	Llave de UPA
CNAV_5_PRODUCTO RES	SI	CONTROL	¿Código del productor (control)?
CNAV_6_UPA_PROD	SI	KEY_COFINCAUPA	Llave Productor
CNAV_6_UPA_PROD	SI	CONTROL	49. ¿Código del productor (control)?
CNAV_7_MAQUINARIA	SI	CONTROL	Llave Maquinaria
CNAV_8_USUELO	SI	KEY_COFINCA	Llave de la finca

Fuente: DANE- 5° CNAM

Dentro de las variables que se consultan en el 5.° CNAM hay tres que se refieren al tamaño de la finca y el cultivo y no solo a la parte de la UPA, estas variables son: área total de la finca, área sembrada y área cosechada. En algunos casos, esta variable podría permitir fácilmente la identificación de la fuente de información.

Las Variables para categorizar son:

- Área total de la finca (ha)
- Área sembrada (ha)
- Área cosechada (ha)
- Área perdida (ha)

Con el fin de permitir el análisis por tamaño de la finca, salvaguardando la reserva estadística, para las variables área total de la finca, área sembrada y área cosechada, se transforma de variable continua a una variable categórica de la siguiente forma:

Tabla 5. Categorización Área.

Tamaño reportado en ha.	Categoría
A < 1	1
1 <= A < 3 ha	2
3 <= A < 5 ha	3
5 <= A < 10 ha	4
10 <= A < 20 ha	5
20 <= A < 40 ha	6
40 <= A < 50 ha	7
50 <= A < 85 ha	8
85 <= A < 100 ha	9
100 <= A < 200 ha	10
200 <= A < 500 ha	11
A >= 500	12

FUENTE: DANE - DANE- 5° CNAM

Tabla 6. Variables para categorizar por área.

Id. tabla	Nombre de la tabla	Id. de la variable	Descripción de la variable
CNAV_1	CNAV_1_FINCAS	AREA	Área total de la finca (ha)
CNAV_2	CNAV_2_UPAS	PERCOS_AREA	¿Cuál fue el área que se perdió?
CNAV_3	CNAV_3_SIEMBRAS	AREA_SEMBRADA	Área sembrada(ha)
CNAV_4	CNAV_4_COSECHAS	AREA_COSECHADA	Área cosechada (ha)
CNAV_8	CNAV_8_USUELO	AREA_ARROZ	Área en arroz
CNAV_8	CNAV_8_USUELO	AREA_MAIZ	Área maíz ha
CNAV_8	CNAV_8_USUELO	AREA_SOYA	Área soya ha
CNAV_8	CNAV_8_USUELO	AREA_ALG	Área algodón ha
CNAV_8	CNAV_8_USUELO	AREA_CULTRAN_OTC	Área otros cultivos transitorios ha
CNAV_8	CNAV_8_USUELO	AREA_PERM	Cultivos permanentes
CNAV_8	CNAV_8_USUELO	AREA_PASFOR	Pastos y Forrajes
CNAV_8	CNAV_8_USUELO	AREA_BAR	Barbecho
CNAV_8	CNAV_8_USUELO	AREA_MALRAS	Maleza y rastrojo
CNAV_8	CNAV_8_USUELO	AREA_BSQNAT	Bosque natural
CNAV_8	CNAV_8_USUELO	AREA_PLFOR	plantación forestal
CNAV_8	CNAV_8_USUELO	AREA_INU	Área INUNDADA en la finca
CNAV_8	CNAV_8_USUELO	AREA_OTACT	Otra actividad
CNAV_4	CNAV_4_COSECHAS	AR_GRANEL	¿Del área cosechada, cuántas fueron recolectadas a partir de los siguientes métodos? (ha) - Granel
CNAV_4	CNAV_4_COSECHAS	AR_BULTO	¿Del área cosechada, cuántas fueron recolectadas a partir de los siguientes métodos? (ha) - Bulto

FUENTE: DANE - DANE- 5° CNAM

La variable cantidad de maquinaria agrícola de la tabla CNAV_7_MAQUINARIA, se recategoriza para evitar la identificación de los productores de lista, como se muestra a continuación.

Tabla 7. Categorización de cantidad de maquinaria.

Cantidad de máquinas	categoría
$C = 1$	1
$2 \leq C \leq 3$	2
$C \geq 4$	3

FUENTE: DANE - DANE- 5° CNAM

4. 1. 4. DISPOSICIÓN DE DATOS

Una vez surtido el proceso mencionado anteriormente, por el cual se preserva la reserva estadística, para las dos estructuras de base de datos del 5° CNAM que serán de consulta en sitio especializado, se tendrá un anexo que contiene el diccionario de variables que son disponibles en la base de datos, con lo cual se facilita su uso al público en general. Adicionalmente, se cuenta con el metadato que describe las características de la operación estadística.

4. 1. 5. APLICACIÓN DE LINEAMIENTOS PARA ANONIMIZACIÓN

A partir de este manual y siguiendo el protocolo establecido por el DANE la Oficina de Sistemas ejecuta las indicaciones dadas en el documento para la generación de las bases de datos anonimizadas y entrega a las instancias correspondientes para que sea consultada por el público.

5. BIBLIOGRAFIA

Diccionario de datos del Quinto Censo Nacional de Arroz 2023
Cuestionario 5° CNAM 2023, Metodología VCNA.