



GOBIERNO
DE COLOMBIA



DANE
INFORMACIÓN
ESTRATÉGICA

Dirección de Regulación, Planeación, Estandarización y
Normalización (DIRPEN)

GIT de Calidad Estadística

Informe de Análisis de la calidad y consistencia de la Base de
Datos “Estadísticas de las personas desmovilizadas que
han ingresado al proceso de reintegración”

Código de Evaluación: 18-PE-012



Octubre de 2018

CONTENIDO

INTRODUCCIÓN.....	3
1. DESCRIPCIÓN DE LA EVALUACIÓN	5
1.1 Desarrollo del proceso de evaluación de la calidad de la base de datos.....	5
2. REVISIÓN DE LA NORMA TÉCNICA PE 1000:2017- REQUISITOS DE BASE DE DATOS	7
3. RESULTADO DEL PROCESO DE VALIDACIÓN DE INTEGRIDAD Y CONSISTENCIA DE LA BASE DE DATOS	12
4. CONCLUSIONES.....	17

INTRODUCCIÓN

El Departamento Administrativo Nacional de Estadística (DANE), como ente rector y coordinador del Sistema Estadístico Nacional (SEN), en cumplimiento del artículo 160 de la Ley 1753 de 2015 y del Decreto 1743 de 2016, define lineamientos, estándares y normas técnicas para la producción, difusión y evaluación de las estadísticas oficiales y para el aprovechamiento estadístico de los registros administrativos.

El esquema para la evaluación de la calidad está basado en los Principios Fundamentales de las Estadísticas Oficiales promulgados por las Naciones Unidas y atributos de calidad adoptados y/o adaptados por el DANE. Para verificar el cumplimiento de estos atributos por parte de las operaciones estadísticas, el DANE estableció los requisitos de calidad que sirven como criterios de medición de la calidad estadística y que están consignados en la Norma Técnica para la calidad del proceso estadístico NTC PE 1000:2017.

Esta norma establece requisitos para las cinco fases del proceso estadístico: 1) Detección y análisis de requerimientos; 2) Diseño y pruebas; 3) Ejecución; 4) Análisis; y 5) Difusión; y tres componentes adicionales relacionados con la Alta Dirección (requisitos generales), la evaluación del desempeño y el proceso de mejora continua. En particular, la evaluación de los requisitos sobre proceso estadístico implica entre otras cosas la revisión de los archivos de datos que están compuestos por microdatos (registros individuales de una operación estadística en una base de datos), los cuales son el resultado de la recolección que se lleva a cabo en el desarrollo de la operación. Sin embargo, las características que presentan los datos en las bases de datos impactan todas las fases del proceso estadístico e incluso a la información estadística, la cual puede llegar a ser incoherente, imprecisa, inexacta o no comparable.

En tal sentido, el equipo evaluador conformado por una Comisión de Expertos Independientes -CEI- (constituida por tres expertos evaluadores externos) y un grupo base interdisciplinario, desarrolla e implementa técnicas de evaluación basadas en las reglas de validación, que establecen los responsables del proceso estadístico, sobre los atributos o variables de la operación estadística conforme a las características o dimensiones de calidad¹ del estándar internacional del Modelo de Calidad de Datos (Norma ISO/IEC 25012)².

Con el uso de estas técnicas y herramientas avanzadas en el diagnóstico de calidad de datos se verifica el mecanismo de validación de integridad y consistencia de la base de datos implementado por los responsables de la operación estadística, lo cual permite determinar el grado de la calidad de los datos y tomar decisiones sobre la gestión de los mismos y por ende impactar positivamente en la exactitud y precisión de los resultados estadísticos finales.

¹ El estándar ISO/IEC 25012 está compuesto por 15 características o dimensiones de calidad (exactitud, completitud, consistencia, credibilidad, actualidad, accesibilidad, confidencialidad, conformidad, eficiencia, precisión, trazabilidad, comprensibilidad, disponibilidad, portabilidad y recuperabilidad).

² La norma ISO/IEC 25012 (Data Quality Model), define un modelo general para la calidad de los datos, aplicable a aquellos datos que se encuentran almacenados de manera estructurada y forman parte de un sistema de información; promovida por el Ministerio de Tecnologías de la Información y las Comunicaciones - MINTIC como estándar de industria. Estándares de Industria, código M_INF_22 p. 19: recuperado de: http://www.mintic.gov.co/arquitecturati/630/propertyvalues-8158_descargable_2.pdf

De esta manera, este informe tiene como objetivo presentar los resultados de la evaluación de la calidad de las bases de datos y su entorno tecnológico realizada en el año 2018 al proceso estadístico de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración” producido por la Agencia para la Reincorporación y la Normalización (ARN). El documento se desarrolla en cuatro capítulos: el primero presenta la descripción del método y la relación de los componentes utilizados para llevar a cabo la evaluación de base de datos; el segundo presenta la descripción de las principales características que se identificaron sobre el componente de base de datos y que se evidencian en las diferentes fases del proceso estadístico de la operación; el tercero contiene el resultado del proceso de validación de integridad y consistencia de la base de datos y el cuarto contiene las conclusiones generales de la evaluación de la calidad de la base de datos.

1. DESCRIPCIÓN DE LA EVALUACIÓN

A continuación, se presenta de forma resumida los principales elementos considerados en la evaluación de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”.

Fecha Inicio Evaluación BD	16/04/2018	Fecha Cierre BD	10/08/2018
Objetivo de la Evaluación			
Evaluar la conformidad del proceso estadístico adelantado por el DANE, en el desarrollo de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, basados en los requisitos establecidos en la Norma Técnica de la Calidad del Proceso Estadístico – Requisitos de Calidad para la generación de Estadísticas – NTC PE 1000:2017.			
Alcance			
La evaluación aplica para el proceso estadístico de las Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración.			
Nombre de la OE			
Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración.			
Criterio de la Evaluación			
Norma Técnica de la Calidad del Proceso Estadístico. Requisitos de calidad para la generación de estadísticas (NTC PE 1000: 2017).			
Periodo evaluado de la base de datos			
Base completa desde el mes de mayo de 2001 hasta el mes de febrero de 2018			

Fuente: DIRPEN – DANE.

1.1 Desarrollo del proceso de evaluación de la calidad de la base de datos

La evaluación de la calidad de la base de datos y su entorno tecnológico involucra todas las fases del proceso estadístico en particular a las actividades desarrolladas para la captura, procesamiento, seguridad, recuperabilidad y confidencialidad de la información. Como se observa en la ilustración 1, para la verificación de la calidad de los archivos de datos se tienen algunos requisitos correspondientes a los capítulos de requisitos generales, diseño y pruebas y ejecución de la NTC PE 1000:2017.

Para hablar de calidad estadística de la base de datos, se debe analizar la calidad de los microdatos de una operación estadística, sobre los cuales se aplica el Modelo de Calidad de Datos (Norma ISO/IEC 25012) para verificar las validaciones que se apliquen a cada variable para que las cifras, los indicadores o los cuadros de salida tengan ese grado de calidad heredado de la calidad de la fuente

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

primaria llamada microdato. Para la evaluación de la calidad de los datos, se selecciona un periodo de referencia de los microdatos de la operación estadística.

La evaluación de la calidad de la base de datos está enmarcada particularmente en los procesos de diseño y ejecución de la operación estadística, cuando en la norma se indica que:

- *NTC PE 1000:2017, Numeral 7.4.3: “La entidad debe **implementar y garantizar la aplicación de las reglas de validación y consistencia** a la información recolectada.”*
- *NTC PE 1000:2017, Numeral 7.2.1.2: “La entidad debe consolidar en una base de datos la información recolectada llevando a cabo:*
 - **Validaciones para revisar, la consistencia de los datos, y**
 - *Los controles sobre la posible pérdida de información.”*

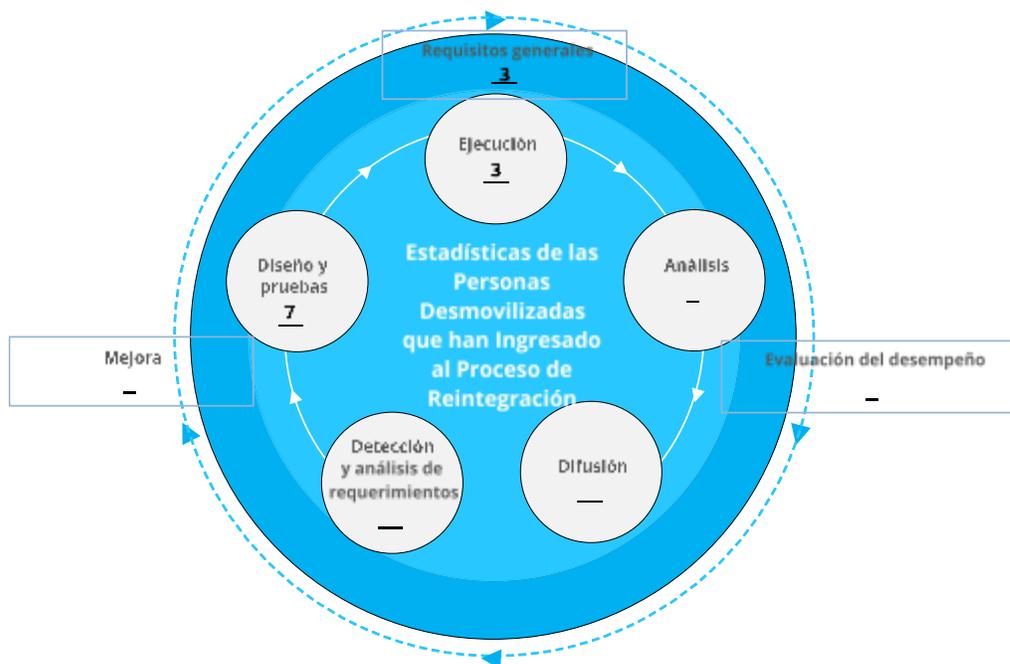
Por lo anterior para la evaluación de la calidad de los archivos de datos se realizan las siguientes actividades por parte del analista de base de datos:

- Los procesos de cargue de los archivos de datos al repositorio de calidad del DANE.
- Revisión y análisis de los componentes del diccionario de datos, entregado por la entidad responsable de la operación estadística en el formato de identificación de evidencias.
- Revisión y análisis de las reglas de validación que aplican sobre cada variable o atributo de la operación estadística a evaluar, entregado por la entidad responsable en el formato de identificación de evidencias.
- Revisión y análisis de la evidencia documental relacionada con los aspectos de captura, procesamiento, seguridad, recuperabilidad y confidencialidad de la información.
- Revisión y análisis de los archivos de datos (microdatos) que componen la operación estadística a evaluar.
- Traducción informática de las reglas de validación a un lenguaje estructurado de consulta SQL.
- Procesamiento de las reglas de validación.
- Generación de los indicadores, que miden el grado de calidad de la información procesada de la operación estadística.
- Generación del detalle de las inconsistencias encontradas (registros inconsistentes).
- Por último, cuantificar y cualificar el grado de calidad de los archivos de datos procesados de la operación estadística.

Con el resultado de estos indicadores se identifican plenamente las inconsistencias de un atributo o variable, de tal manera que se pueda determinar:

- ¿Cuáles son las variables afectadas?
- ¿Cuáles son las de mayores incidencias con respecto a todas las variables procesadas de uno o todos los archivos de datos?
- ¿Cuáles son las inconsistencias más comunes?
- Identificar plenamente y en detalle, ¿Cuáles son las inconsistencias?
- ¿Cuáles son las inconsistencias encontradas (reglas que no cumplen) al aplicar las reglas de validación?

Ilustración 1. Número de requisitos evaluados por componente de la NTC PE 1000: 2017, relacionados con la base de datos



Fuente: DIRPEN – DANE.

2. REVISIÓN DE LA NORMA TÉCNICA PE 1000:2017- REQUISITOS DE BASE DE DATOS

En esta sección se presenta una breve descripción de las principales características que se identificaron en la evaluación de calidad de la base de datos, tanto en la fase de revisión documental como en la fase de revisión en sitio, sobre el componente de base de datos y que se evidencian en las diferentes fases del proceso estadístico de la operación “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”. Dado que para la consolidación de la base de datos los responsables de la operación realizan actividades que involucran a la Alta Dirección y que se desarrollan en el diseño y pruebas, la recolección y el procesamiento. A continuación, se relacionan las prácticas identificadas para cada uno de los capítulos de la NORMA TÉCNICA PE 1000:2017 en los cuales se identifican requisitos relacionados con la consolidación de la base de datos.

Requisitos Generales

La Agencia para la Reincorporación y la Normalización – ARN, cuenta con mecanismos para garantizar que la información de las fuentes se conserva de manera confidencial y segura, dado que disponen del documento “Manual del Sistema de Gestión de Seguridad de la Información”, el cual es

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

muy completo y detallado sobre todas las políticas de seguridad de la información que implementa la agencia y el cual es conocido por todos los colaboradores de la ARN, sin importar su tipo de vinculación, donde se informa como se deben aplicar las políticas, normas y procedimientos en materia de seguridad de la información; de tal forma que se desarrollan capacitaciones para dar a conocerlas, además dentro de los contratos de personal, existe una cláusula en la cual se debe conocer el Manual del Sistema de Gestión de Seguridad de la Información; así mismo, dentro de los contratos de prestación de servicios de personal y convenios con otras entidades, se evidenció el manejo de cláusulas de confidencialidad de la información, donde se fijan los términos y condiciones bajo los cuales se debe mantener la confidencialidad de la información y la máxima protección de los datos que se conozca con ocasión del contrato. Adicionalmente el personal de la Agencia firma un documento llamado “Acta de Compromiso y Autorización sobre Confidencialidad y Manejo de la Información”. Otro de los mecanismos utilizados por la ARN para garantizar la confidencialidad de las personas desmovilizadas que ingresan al proceso de reintegración, es difundir los resultados de manera agregada a nivel nacional y por departamento de las diferentes situaciones, temáticas o dimensiones que se presentan en el proceso de reintegración de los desmovilizados.

Con respecto a la seguridad de la información recolectada, para prevenir el acceso, uso no deseado, daños o pérdida fortuita de la información de la operación estadística, la Agencia para la Reincorporación y la Normalización – ARN, cuenta con mecanismos documentados para generar copias de respaldo diario incremental, semanal y mensual y restauración de la información, lo cual se lleva a la práctica y pudo ser verificado en la visita de sitio; además, cuenta con servidores alternos en Bogotá y Medellín.

Por otro lado los usuarios de la Entidad sólo pueden acceder a aquellos sistemas de información a los que estén autorizados y que sean necesarios para el desempeño de sus actividades, cumpliendo con responsabilidades tales como: resguardar la confidencialidad de la información a la cual tiene acceso, usar de forma adecuada los Sistemas de información con sus respectivos procedimientos, mecanismos, controles de identificación y autenticación y utilizar las contraseñas de forma adecuada, entre otras.

La Agencia para la Reincorporación y la Normalización, cuenta con manuales y formatos para el manejo e implementación de la seguridad, confidencialidad y respaldo de la información, los cuales se relacionan a continuación:

- MANUAL DEL SISTEMA DE GESTIÓN DE SEGURIDAD INFORMACIÓN
- Formato: Acta de Compromiso y Autorización sobre Confidencialidad y Manejo de la Información” Manual del sistema de gestión de seguridad información Política y objetivos de seguridad

La operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, dispone de una infraestructura tecnológica para la demanda del almacenamiento y procesamiento oportuno de la información; su herramienta de software es Microsoft Dynamics CRM, que facilita el acceso a la información en tiempo real y garantiza su seguridad; bajo esta plataforma se desarrolla el Sistema de Información para la Reintegración (SIR), sobre el cual se tienen implementados los módulos y funcionalidades.

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

Para lograr los objetivos y el alcance temático, la operación estadística dispone de módulos de recolección de información y de instructivos de diligenciamiento, tales como: (línea base de reintegración - Caracterización del grupo familiar - Encuesta de Hábitat - Instrumento de tasa de desocupación - Instrumento de Seguimiento a Unidades de Negocio).

El diligenciamiento digital cuenta con diferentes aplicativos, uno de ellos se relaciona con el registro de la información directamente desde el SIR, los otros dos son los aplicativos ARPA y SARA. ARPA es un desarrollo de Microsoft donde se registra información del proceso de reintegración y de los instrumentos de monitoreo y seguimiento sin tener una conexión a internet; mientras que SARA es un ambiente web, donde se registra información de una forma más amigable con el usuario, pero requiere una conexión a internet. Así mismo la entidad cuenta con el portal SIR, una plataforma en la que se pone a disposición de los usuarios reportes, videos explicativos y otro material de apoyo.

Por otro lado, tanto para la integración y procesamiento de los datos almacenados, se cuenta con el Datamart, la cual es una bodega de datos, con esta herramienta se actualiza la información de manera automática, diariamente y se genera la Base de datos Nacional, la cual es empleada para dar respuesta a las diferentes necesidades de información de los usuarios internos y externos.

Diseño y pruebas

Los responsables de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, han establecido las reglas de validación y consistencia, las cuales se encuentran descritas en el documento de especificaciones y consistencia.

Así mismo, los procesos de recolección y procesamiento de la información están definidos en el documento metodológico y manuales de diligenciamiento que contiene las pautas y conceptos requeridos para el correcto diligenciamiento de la información de cada componente de la operación estadística, los cuales son acordes con los objetivos de la investigación.

A continuación, se relaciona la documentación técnica disponible de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”:

- Manual del sistema de gestión de seguridad de la información
- Manual de consistencia y validación de las variables de la base nacional de participantes que ingresaron al proceso de reintegración
- Instructivo descripción de variables de la base de datos nacional
- Instructivo de insumos empleados para la consolidación de la información en la base de datos nacional
- Identificación evidencias 12042018.xls - Hoja DICCIONARIO DATOS
- Metodología de las Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración
- Instructivos de diligenciamiento: (Línea base de reintegración / Caracterización del grupo familiar / Encuesta de Hábitat / Instrumento de tasa de desocupación / Instrumento de Seguimiento a Unidades de Negocio)
- Manuales de usuario SIR (20 componentes o módulos)
- Manual de Usuario DATAMART

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

- INSTRUMENTOS DE RECOLECCIÓN DE INFORMACIÓN: (Encuesta Tasa de Desocupación / IR-F-11 Instrumento de seguimiento a unidades de negocio (ISUN) / Plantilla censo de familia de las PPR / PLANTILLA CENSO DE HÁBITAT / Plantilla instrumento línea base
- MANUALES DE DILIGENCIAMIENTO: (Documento apoyo aplicación censo familia / Documento de apoyo para la aplicación de la encuesta de hábitat / Instructivo para el diligenciamiento de la línea base a las personas en proceso de reintegración / IR-G-08 Guía de Aplicación del Instrumento (ISUN)
- Modelo Entidad Relación

Con relación a la infraestructura tecnológica, la operación estadística dispone de tres sistemas de información para la captura de los datos, el SIR, SARA y ARPA, una vez la información es registrada y depurada en su totalidad, esta información pasa a la bodega de datos DATAMART, herramienta que permite generar los productos para responder a las necesidades de información de los usuarios internos y externos de la entidad, cumpliendo con directrices, lineamientos de seguridad y protección de la información, a través de la gestión segura de los activos de información de acuerdo a lo establecido en el manual del Sistema de Gestión de Seguridad de la Información.

Ejecución

En esta fase se desarrollan las actividades definidas en la fase de diseño y pruebas, las cuales se encuentran descritas en el documento metodológico, según el cual, el proceso de recolección es realizado por un equipo de profesionales llamados reintegradores, los cuales se encuentran ubicados ya sea en los grupos territoriales y puntos de atención que tiene la Agencia para la Reincorporación y la Normalización, quienes tienen contacto directo con las personas en proceso de reintegración y registran la información de las personas desmovilizadas en el Sistema de Información para la Reintegración – SIR. Por otro lado, los técnicos tienen dentro de sus actividades el registro en el SIR de todas las atenciones, aperturas, seguimientos, cierres de los casos, novedades y la actualización permanente en el Sistema de los datos de ubicación y demás datos de contacto de los desmovilizados en proceso de reintegración. Es importante precisar que el desmovilizado que ingresa al proceso de reintegración debe surtir unas fases las cuales se llaman dimensiones y estas manejan una periodicidad de recolección diferente, por lo cual la recolección del desmovilizado puede llegar a durar seis años y medio, que es aproximadamente lo que puede durar el programa de reintegración para los desmovilizados. Dado que el sistema de información se encuentra en un ambiente WEB, la información es registrada, actualizada y almacenada, en tiempo real, en una base de datos única y centralizada, además es importante precisar que para el procesamiento de la información, como mecanismo para controlar la calidad de la información, el SIR genera alertas para los usuarios a medida que van diligenciando los datos, en donde el sistema arroja mensajes de advertencia sobre completitud de los datos en las variables que no deben quedar vacías, datos mal registrados, entre otros; por lo que se evidenció, que se encuentran implementadas las validaciones registradas en los documentos relacionados con la consistencia y validación de las variables de ingreso a la base de datos.

Frente al procesamiento de la información, la Agencia para la Reincorporación y la Normalización, cuenta con la infraestructura tecnológica para almacenar y procesar la información, una vez registrada la información de la persona desmovilizada para el proceso de reintegración, pasa la

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

información al DATAMART, una bodega de datos que ha permitido optimizar el procesamiento de la información. Esta herramienta actualiza la información de manera automática y diariamente. A partir de esta información, la agencia genera la Base de datos Nacional la cual es empleada para dar respuesta a las diferentes necesidades de información de los usuarios internos y externos.

Por otro lado el mismo sistema de información genera los cuadros de salida, los cuales son actualizados periódicamente y están disponibles para ser consultados y descargados en la página web de la agencia.

Con respecto a los datos de la operación estadística, se evidenció su integridad, dado que a pesar de tener varios aplicativos para la recopilación de los datos, toda la información es transferida a un mismo repositorio de base de datos, desde donde finalmente son generados los cuadros para su publicación.

La operación estadística, tiene documentada las actividades desarrolladas durante la consolidación y procesamiento de la información, por tanto, que en la base de datos queda registrada toda la información correspondiente y toda la trazabilidad de los cambios que se hayan realizado a los datos, dado que el sistema de información tiene implementado un log de auditoría, además la entidad cuenta con la aplicación de ARANDA, donde se registran, almacenan y centralizan las solicitudes de soporte tecnológico de los usuarios, a través de esta aplicación se asignan las incidencias a los especialistas para ser solucionadas las inquietudes que surjan durante el proceso de recopilación y sistematización de la información; de igual manera, por medio de esta aplicación se realizan solicitudes de ajustes a la recolección de información.

Para evitar la pérdida de información y protegerla contra modificación o alteración, se evidenció en la visita en sitio que la oficina de tecnologías de la información, realiza copias de seguridad de los servidores donde reposa la información de la agencia, incluyendo la base de datos de la operación estadística, para lo cual tienen automatizadas unas rutinas para que se realice cuatro (4) backups diarios incrementales, backup semanal y backup mensual, con el propósito de evitar que se ponga en riesgo la disponibilidad del sistema de información. Así mismo estas copias de seguridad son resguardadas en cintas y custodiadas por una empresa externa a la agencia, con la cual se tiene contrato. De igual forma, se realizan pruebas de restauración de backup, con el propósito de verificar que efectivamente las copias de seguridad se están realizando adecuadamente.

Para mantener la disponibilidad de la operación estadística, la ARN, cuenta con un servidor alternativo en la ciudad de Medellín, es decir un servidor que contiene una réplica del servidor principal donde se encuentra la base de datos de la operación, con el propósito de facilitar el acceso a la información aun cuando haya fallos en el servicio del servidor principal.

Todos los usuarios de la agencia sólo pueden acceder a aquellos sistemas de información a los que estén autorizados y que sean necesarios para el desempeño de sus actividades, además las credenciales de acceso a la red y/o recursos informáticos (Usuario y Clave) son de carácter estrictamente personal e intransferible, los colaboradores no deben revelar estas a terceros, ni utilizar claves ajenas, así mismo, cada usuario es responsable de los registros y/o modificaciones de información que se hagan a nombre de su cuenta de usuario, toda vez que la clave de acceso es de carácter personal e intransferible.

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

3. RESULTADO DEL PROCESO DE VALIDACIÓN DE INTEGRIDAD Y CONSISTENCIA DE LA BASE DE DATOS

Este capítulo contiene los resultados del proceso de validación de integridad y consistencia de la base de datos. En primer lugar, se relacionan los archivos de base de datos suministrados por el equipo responsable de la operación estadística, en segundo lugar, se presentan las inconsistencias identificadas a través de la aplicación de las reglas de validación y consistencia definidas por los responsables de la operación; así como el detalle de las inconsistencias encontradas (a nivel de registro, campo y variable) y por último, se presentan los indicadores generados para la medición del grado de calidad de la información procesada, cuyo objetivo es cuantificar y cualificar el grado de calidad de los archivos de datos procesados.

Este capítulo contiene los resultados del proceso de validación de integridad y consistencia de la base de datos. En primer lugar, se relacionan los archivos de base de datos suministrados por el equipo responsable de las “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, en segundo lugar, se presentan las inconsistencias identificadas a través de la aplicación de las reglas de validación y consistencia definidas por los responsables de la operación; así como el detalle de las inconsistencias encontradas (a nivel de registro, campo y variable) y por último, se presentan los indicadores generados para la medición del grado de calidad de la información procesada de la operación estadística, cuyo objetivo es cuantificar y cualificar el grado de calidad de los archivos de datos procesados.

3.1. Relación y descripción de archivos de base de datos recibidos

A continuación, se relacionan las evidencias documentales y los archivos de datos que fueron suministrados por los responsables de la operación estadística, y que fueron el insumo para la revisión del componente de base de datos.

Tabla 1: Relación evidencias documentales suministradas por la entidad

Documentación de la Base de Datos
<ul style="list-style-type: none">Manual del sistema de gestión de seguridad de la informaciónManual de consistencia y validación de las variables de la base nacional de participantes que ingresaron al proceso de reintegraciónInstructivo descripción de variables de la base de datos nacionalInstructivo de insumos empleados para la consolidación de la información en la base de datos nacionalIdentificación evidencias 12042018.xls - Hoja DICCIONARIO DATOSMetodología de las Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegraciónInstructivos de diligenciamiento: (Línea base de reintegración / Caracterización del grupo familiar / Encuesta de Hábitat / Instrumento de tasa de desocupación / Instrumento de Seguimiento a Unidades de Negocio)Manuales de usuario SIR (20 componentes o módulos)Manual de Usuario DATAMARTINSTRUMENTOS DE RECOLECCIÓN DE INFORMACIÓN: (Encuesta Tasa de Desocupación / IR-F-11 Instrumento de seguimiento a unidades de negocio (ISUN) / Plantilla censo de familia de las PPR / PLANTILLA CENSO DE HÁBITAT / Plantilla instrumento línea baseMANUALES DE DILIGENCIAMIENTO: (Documento apoyo aplicación censo familia / Documento de apoyo para la aplicación de la encuesta de hábitat / Instructivo para el diligenciamiento de la línea base a las personas en proceso de reintegración / IR-G-08 Guía de Aplicación del Instrumento (ISUN)Modelo Entidad Relación

Fuente: DIRPEN – DANE.

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

A continuación, se relacionan los archivos de datos entregados por el responsable de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, para carga y evaluación.

Tabla 2. Archivos de base de datos recibidos, cargados y evaluados

Nombre Archivo Bases de Datos	Descripción Archivo Bases de Datos	Archivo cargado	Total registros
Base Nal Febrero 2018	Caracterización de las personas desmovilizadas que han ingresado al proceso de reintegración	ARN_18_OPERACIÓN ESTADÍSTICA	51.253
Total de registros			51.253

Fuente: Dirpen– DANE.

A partir de los archivos cargados se obtiene la siguiente ficha técnica donde se describe cuantitativamente el contenido de la base de datos.

Tabla 3. Ficha técnica evaluación de la base de datos

Período de la base de datos evaluada	Completa a febrero de 2018
Archivos procesados	1
Registros procesados	51.253
Campos procesados	5.022.794
Reglas de validación aplicadas	210
Variables analizadas	100
Cantidad de procesamientos de la base de datos	2

Fuente: DIRPEN- DANE

3.2. Resultados de la verificación de la implementación de las reglas de validación y consistencia

En esta sección se presentan las inconsistencias identificadas a través de la aplicación de las reglas de validación y consistencia definidas por los responsables de la operación Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración, a continuación, se presenta el resumen de las inconsistencias encontradas.

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración
Código de Evaluación: 18-PE-012

Tabla 4. Inconsistencias presentadas en la aplicación de reglas de validación

	ANALIZADAS	AFFECTADAS
VARIABLES	100	2
REGISTROS	51.253	1.319
CAMPOS	5.022.794	2.568

Fuente: Dirpen – DANE.

Como se puede observar en la anterior tabla, se identificaron inconsistencias en dos (2) variables, las cuales son detalladas en la siguiente tabla.

Tabla 5. Principales hallazgos evaluación de la base de datos

TABLA: DANE_18_OPERACIÓN ESTADÍSTICA_PRODC
<p>Variable: MUNICIPIO; Municipio de residencia del individuo al mes de corte. Regla: RV150 - 122: Reglas de integridad referencial 1. se verifica que los códigos de los municipios se encuentren en la DIVIPOLA Se detectaron 1.319 campos inconsistentes en la variable</p>
TABLA: DANE_18_OPERACIÓN ESTADÍSTICA_PRODC
<p>Variable: DEPARTAMENTO; Departamento de residencia del individuo al mes de corte. Regla: RV400: Reglas de dominio 1. se verifica que los códigos de los departamentos se encuentren en la DIVIPOLA Se detectaron 1.249 campos inconsistentes en la variable.</p>

Fuente: Dirpen – DANE.

Adjunto a este informe se entrega el Anexo 2, que contiene la relación detallada de los registros que presentaron una o varias inconsistencias durante la evaluación, el cual se denomina ERR_ARN_18_OPERACIÓN ESTADÍSTICA_DATA_TABLE.

3.3 Indicadores de calidad

Los indicadores de calidad que a continuación se describen, corresponden a la medición de la calidad de los datos de uno o un conjunto de archivos evaluados, calculados mediante un método de medición directa, que posteriormente son ponderados para un cálculo integral que permita determinar el grado de calidad total de los archivos, variables y campos procesados mediante la aplicación de reglas de validación sobre sus atributos o variables.

Con base en estas consideraciones, se definen y calculan los siguientes indicadores:

- Indicador de calidad a nivel de registro:** indica el número de registros de cada tabla o archivo de datos procesados, que se vieron afectados o son inconsistentes al aplicar las reglas de validación.

Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

ARCHIVO	REGISTROS	REGISTROS INCONSISTENTES	INDICADOR DE CALIDAD
ARN_18_OPERACIÓN ESTADÍSTICA	51.253	1.319	97,43%
TOTAL	51.253	1.319	97,43%

Fuente: DIRPEN – DANE.

b. Indicador de calidad a nivel de campo: indica el número de campos (valor de la variable o atributo) de una tabla o archivo de datos procesados, que se vieron afectados o son inconsistentes al aplicar las reglas de validación. El detalle de cada una de las inconsistencias se entrega en el archivo anexo: RESUMEN_INCONSISTENCIAS_23-05-2018.xls.

ARCHIVO	CAMPOS	CAMPOS INCONSISTENTES	INDICADOR DE CALIDAD
ARN_18_OPERACIÓN ESTADÍSTICA	5.022.794	2.568	99,95%
TOTAL	5.022.794	2.568	99,95%

Fuente: DIRPEN – DANE.

c. Indicador de calidad a nivel de variable: indica el número de variables de la operación estadística que se vieron afectadas al aplicar las reglas de validación.

$$ICV = 1 - \left[\frac{\text{Número de variables afectadas}}{\text{Número variables analizadas}} \right]$$

$$ICV = 1 - [2/100]$$

ICV = 98%

d. Indicador total de la calidad de la base de datos: finalmente, para obtener un indicador de calidad total, se ponderan los indicadores individuales a nivel de variable, campo y registro con diferentes pesos, los cuales dependen de su importancia y relevancia para el análisis de calidad como se indica a continuación:

INDICADOR DE CALIDAD	INDICADOR	PONDERACIÓN	(%)
A NIVEL DE CAMPO	99,95%	60	59,97%
A NIVEL DE VARIABLE	98,00%	30	29,40%
A NIVEL DE REGISTRO	97,43%	10	9,74%
Total Promedio Ponderado			99,11%

Fuente: DIRPEN – DANE

Una vez cuantificados los resultados de la base de datos, a través de la obtención del promedio ponderado de la calidad de la base de datos, este valor se ubica de acuerdo a los siguientes criterios cualitativos que fueron dedefinidos en la metodología de la evaluación para calificar la calidad de toda la base de datos (archivos de datos) evaluada:

Informe de Análisis de la calidad y consistencia de la Base de Datos



Estadísticas de las Personas Desmovilizadas que han Ingresado al Proceso de Reintegración

Código de Evaluación: 18-PE-012

Excelente	Buena	Aceptable	Deficiente
Mayor o igual a 95%	Entre 85 % y 94%	Entre 75 % y 84%	Menos del 75%

Fuente: DIRPEN – DANE.

De esta manera, se concluye que siendo el total promedio ponderado de 99,11%, el nivel de calidad de la base de datos de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración” es excelente.

4. CONCLUSIONES

El indicador de calidad integral de toda la base de datos presenta un excelente grado de calidad del archivo de datos procesado con un 99,11%, de acuerdo con los parámetros cuantitativos y cualitativos establecidos; por lo anterior, se puede evidenciar que el aseguramiento de la calidad de los datos aplicada a los microdatos de la operación estadística “Estadísticas de las personas desmovilizadas que han ingresado al proceso de reintegración”, trasciende de manera positiva en la calidad de las estadísticas generadas.

Dentro de la documentación entregada por la Agencia para la Reincorporación y la Normalización, se evidenció la disponibilidad del manual de validación y consistencia y el diccionario de datos, donde se relacionan las variables de ingreso a la base de datos con sus respectivas reglas de validación y consistencia; sin embargo, algunas de las reglas están desactualizadas con respecto a sus dominios y no son claras al no contar con operadores lógicos, situación que podría ocasionar inconvenientes futuros en la actualización o nuevo desarrollo del sistema de información de la operación estadística. Lo anterior puede generar un incumplimiento del a los numerales 6.5.2 y 6.6.2 de la NTC PE 1000:2017

Las reglas de validación y consistencia se encuentran documentadas. Sin embargo, en la validación de la calidad de la base de datos, se evidencian inconsistencias en el uso del estándar de la Codificación de la División Política Administrativa - DIVIPOLA, encontrando valores sustitutos en los campos de municipio y departamento, tales como “No registra” y “27086”, los cuales no hacen parte de la codificación mencionada, situación que puede afectar la comparabilidad en la utilización de los microdatos de la operación y las características de los datos conforme a los estándares, así mismo generar un incumplimiento al numeral 7.4.3 de la NTC PE 1000:2017.

Por otro lado, se destaca que la agencia cuenta con un documento muy detallado sobre las políticas, normas y procedimientos que se deben implementar dentro de la entidad para garantizar la seguridad de la información, lo que se convierte en una fortaleza al contar con el “manual del sistema de gestión de seguridad información”

En la revisión en sitio se identificó, que esta operación estadística dispone de mecanismos para garantizar la confidencialidad de la información, tales como: publicación de resultados agregados, acceso seguro a los datos ingresando a la aplicación con usuario y clave de acceso, con respecto al respaldo de la información se identificó que realizan backups y su restauración, con el propósito de asegurar la continuidad de la operación.

Sobre la infraestructura tecnológica, se evidenció que la agencia cuenta con instalaciones, equipos de cómputo, redes y mobiliario suficiente para el desarrollo del proceso estadístico, facilitando de esta forma los procesos operativos, de análisis y difusión de las estadísticas generadas.